Towards a functional characterization of meiotic recombination in rapeseed : analysis of the meiotic transcriptome and hyper-recombinant mutants

Aurélien Blary

NNT : 2016SACLS576

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème

Spécialité de doctorat : Biologie

Par

Aurélien Blary

Towards a functional characterization of meiotic recombination in rapeseed; analysis of the meiotic transcriptome and hyper-recombinant mutants

Thèse présentée et soutenue à Versailles, le 20 décembre 2016 :

Composition du Jury :

Dr. Shykoff Jacqui	Directeur de recherche	Paris Saclay	Présidente du Jury
Dr. Lashermes Philippe	Directeur de recherche	IRD	Rapporteur
Dr. David Jacques	Enseignant-chercheur	Mtp Supagro	Rapporteur
Dr. Byrne Ed	Project Leader	KWS	Examinateur
Dr. Sourdille Pierre	Directeur de recherche	INRA	Examinateur
Dr. Jenczewski Eric	Directeur de recherche	INRA	Directeur de thèse

Remerciements

Je tiens tout d'abord à remercier les membres du jury qui ont accepté d'évaluer mon travail: les rapporteurs Philippe Lashermes et Jacques David, ainsi que les examinateurs Jacqui Shykoff, Ed Byrne et Pierre Sourdille pour leur lecture attentive du manuscrit et les discussions qui ont suivi. Je remercie aussi les membres de mon comité de thèse, Françoise Budar, Anne-Marie Chèvre, Frédéric Choulet, Olivier Martin et Raphaël Mercier pour la pertinence de leurs remarques, propositions et critiques.

J'adresse plus particulièrement mes remerciements à Eric Jenczewski, pour m'avoir guidé durant les premiers six mois de Master à Versailles, m'avoir préparé au concours de l'école doctorale (avec l'aide de l'ensemble des membres du Bâtiment 7 et plus particulièrement celle de l'équipe «Méiose et Recombinaison» Raphaël, Mathilde, Christine et Fabien) et m'avoir accompagné durant ces trois années de thèse. C'est un réel plaisir de travailler à ses côtés, sa grande rigueur scientifique combinée à ses qualités humaines en font un encadrant exceptionnel.

L'environnement scientifique de qualité et la constante émulation intellectuelle ont beaucoup contribué à rendre ce travail de thèse très enrichissant. Je remercie Mathilde Grelon pour sa relecture critique du manuscrit et ses précieux conseils. Je renouvelle mon chaleureux remerciement à toute l'équipe «Méiose et Recombinaison» pour leur aide technique et leur présence durant ces années de même qu'à l'ensemble des membres du Bâtiment 7.

Toute ma gratitude à Catherine qui m'a pris en charge durant les premières semaines de stage et qui a su si bien s'occuper de nos plantes.

Je voudrais aussi souligner l'implication de l'équipe bio-informatique du centre, Delphine, Joseph et Fabienne, qui m'ont accueilli au sein de leur groupe pendant six mois et qui ont grandement contribué à l'analyse des données. J'ai beaucoup appris à leur contact.

Merci également à Guillem Rigaill qui a été très pédagogue pour m'expliquer certaines analyses statistiques et qui a été à la source de propositions pour les analyses d'expressions.

Je salue l'équipe d'Anne-Marie Chevre à Rennes qui a effectué un travail conséquent et de qualité pour la partie FANCM. Merci aussi à Liudmila pour ses conseils en cytologie.

Je n'oublie pas la direction de notre école doctorale, Jacqui, Marianne et Martine quant au suivi des doctorants et de leur association «Doc'enherbe» dont je salue chaleureusement tous les membres. Enfin, un grand merci à tous les thésards de l'Institut et de l'école doctorale que j'ai eu l'occasion de rencontrer au cours de ces années.

Table of contents

Introduction **1**

Chapter 1: Bibliographic review **4**

1.1 Progression of meiosis, as seen through the prism of chromosome association/segregation....4

1.2 The molecular mechanisms of meiotic recombination....12

1.3 The progression of meiosis is intertwined with meiotic recombination....17

1.4 The patterning of meiotic COs formation 20

1.5 How to deal with the polyploidy situation?....25

1.6 Ploidy level and recombination frequencies....29

1.7 Impact of recombination on genetic diversity 30

1.8 CO frequencies and selection 33

1.9 How to tackle some of the breeder's challenges? 35

Chapter 2: The plant model **40**

2.1 Origin of *Brassica napus*....40

2.2 Relevance for breeding....47

2.3 Meiosis in *Brassica napus*....49

2.4 AAC triploid hybrids....56

Chapter 3: Objectives of the PhD **57**

Chapter 4: Deciphering the main source of variation for the meiotic transcriptome of *B. napus* **58**

4.1 Introduction 58

4.2 Manuscript: Homoeologous exchanges drive extensive dosage dependent changes in gene expression and influence allopolyploid genome evolution..... 61

4.3 Overview of the transcriptome of *Brassica napus* meiocytes 137

4.3.1 Objectives 137

4.3.2 Details on the experimental design 141

4.3.3 Details on the mapping 144

4.3.4 De novo transcriptome assembly 146

4.3.5 Description of the meiotic transcriptome of *B. napus* 148

4.3.6 Partitioning the source of transcriptome variation 151

4.3.7 Differential gene expression between homoeologs does not result in global sub-genome dominance 155

4.3.8 Variation of the transcriptome between *Darmor-bzh* and *Yudal* 157

4.3.9 Ploidy change had limited impact on meiotic gene expression 160

4.3.10 A closer look into *PrBn* confidence interval 162

4.3.11 Conclusions and Perspectives 163

Chapter 5: FANCM **165**

5.1 Manuscript: FANCM limits meiotic COs in *Brassica* crops 165

Chapter 6: General discussion **193**

6.1 The meiotic transcriptome is highly variable within *Brassica napus* 193

6.2 Phenotypic consequences of HEs 196

6.3 The anti-CO activity of FANCM is conserved in the *Brassica* 200

References **203**

Summary **228**

List of figures

Figure 1: Classical breeding schemes 1
Figure 2: Progression of meiosis and cytological manifestation of meiosis stage in plant model species 5
Figure 3: The cohesin complex 7
Figure 4: The Synaptonemal complex 9
Figure 5: Early Double Strand Break (DBS) Repair Process repair process in budding yeast 11
Figure 6: Meiotic recombination mechanisms 13
Figure 7: Circuits that regulate double strand breaks timing and patterning 16
Figure 8: The beam-film model.....19
Figure 9: Number of crossovers (COs) per chromosome per meiosis in a variety of eukaryotes 20
Figure 10: Genomic changes with crop domestication and breeding 31
Figure 11: Collinearity of A and C subgenomes.....41
Figure 12: The course of polyploidy events from ancestral angiosperm to Brassica species 43
Figure 13: Flow chart of the ‘‘two-step theory’’ to explain the genome triplication that occurred in the early stages of the origin of *B. rapa* species 44
Figure 14: Rich morphotypes of *Brassica* plants 46
Figure 15: Meiosis in Euploid *B. napus* 48
Figure 16: Meiosis in Allohaploid *B. napus* 50
Figure 17: Natural variation in CO frequencies in *B. napus* allohaploid.....52
Figure 18: Mapping of *PrBn* and epistatic QTLs 53
Figure 19: Genetic maps of the A7 Linkage Group in Progeny of the Diploid (ArAr), Allotriploid (ArArC), and Allotetraploid (ArArCC) Hybrids.....55
Figure 20: Setting up the experimental design.....139
Figure 21: Experimental design for RNA-seq transcriptomic analysis.....140
Figure 22: Cumulative proportion of reads that map on the reference genome when varying mapping stringency parameter 142

Figure 23: A/C bias as assessed by RNA-seq and pyrosequencing 143
Figure 24: Gene distribution according to total read count.....147
Figure 25: Correlation of meiotic gene expression in *A. thaliana* and *B. napus* 147
Figure 26: Score plot (homoeologous gene pairs).....149
Figure 27: Score plot (entire gene set) 150
Figure 28: Detection of a newly formed HE segregating within *Darmor-bzh* euploids biological replicates.....152
Figure 29: Consistency of total reads counts across samples.....154
Figure 30: Differentially expressed meiotic genes between *Darmor-bzh* and *Yudal*...156
Figure 31: Differentially expressed homoeologous gene pairs between *Darmor-bzh* and *Yudal*.....156
Figure 32: Genes and SNP density on the C09 chromosome 161

List of tables

Table 1: Characterization of de novo transcripts via a BLAST analysis (BLAST2GO) 145
Table 2: Differential expression in *B. napus* transcriptions factors that are within the most up-regulated genes in meiosis in both *Arabidopsis* and maize 158
Table 3: Differential expression for *B. napus* transcriptions factors 158

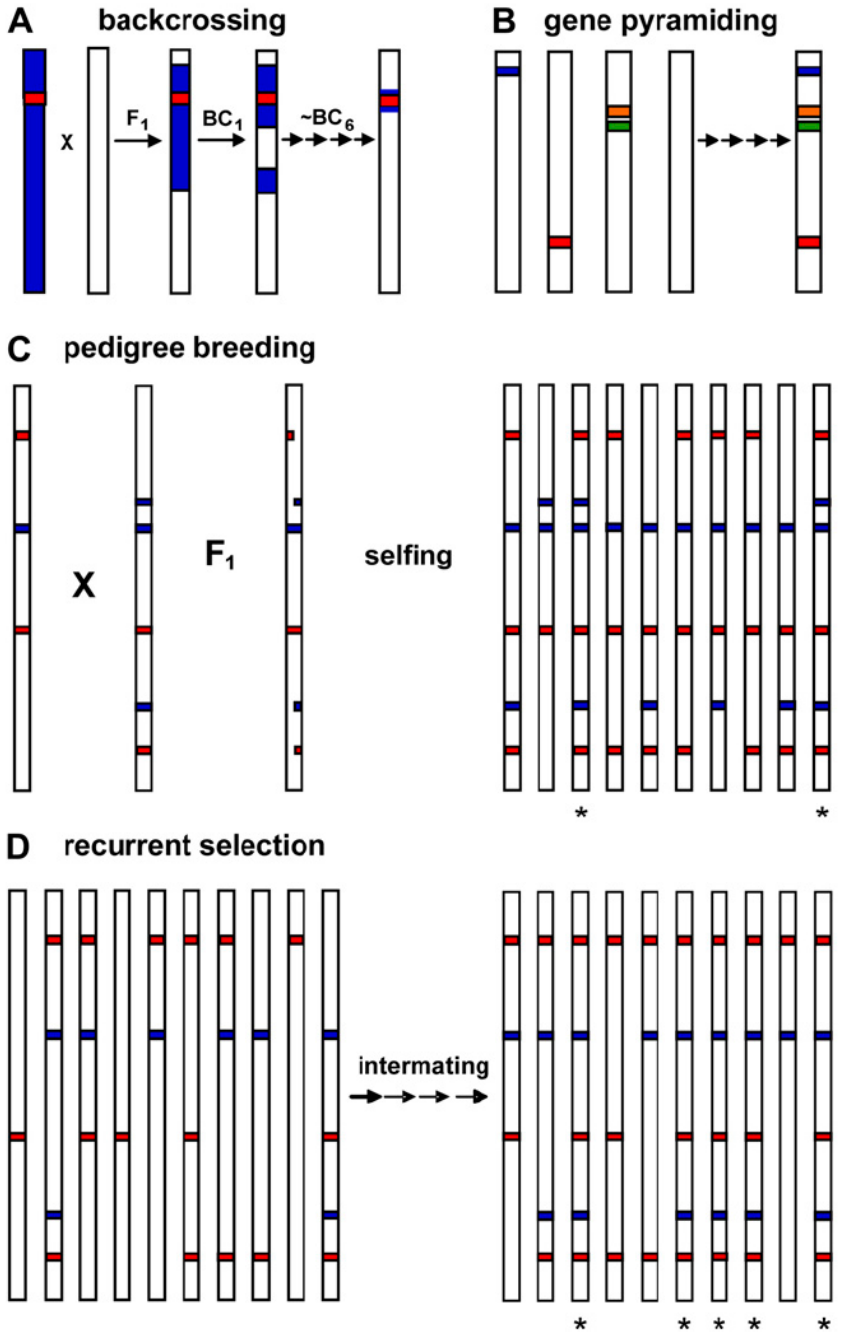


Figure 1: Classical breeding schemes [figure and text reproduced from (Moose and Mumm, 2008)]

Each vertical bar is a graphical representation of the genome for an individual within a breeding population, with colored segments indicating genes and/or QTLs that influence traits under selection. Genes associated with different traits are shown in different colors (e.g. red, blue). ‘‘X’’ indicates a cross between parents, and arrows depict successive crosses of the same type. Asterisk below an individual signifies a desirable genotype.
(A) Backcrossing. A donor line (blue bar) featuring a specific gene of interest (red) is crossed to an elite line targeted for improvement (white bar), with progeny repeatedly backcrossed to the elite line. Each backcross cycle involves selection for the gene of interest and recovery of increased proportion of elite line genome.
(B) Gene pyramiding. Genes/QTLs associated with different beneficial traits (blue, red, orange, green) are combined into the same genotype via crossing and selection.
(C) Pedigree breeding. Two individuals with desirable and complementary phenotypes are crossed; F1 progeny are self-pollinated to fix new, improved genotype combinations.
(D) Recurrent selection. A population of individuals (10 in this example) segregate for two traits (red, blue), each of which is influenced by two major favorable QTLs. Intermating among individuals and selection for desirable phenotypes/genotypes increases the frequencies of favorable alleles at each locus. For this example, no individual in the initial population had all of the favorable alleles, but after recurrent selection half of the population possesses the desired genotype.

Introduction

The objective of commercial plant breeding is the production of new varieties with superior agronomical traits compared to what is already present on the market. Traits are encoded by genes, which exist in different versions, termed alleles, on chromosomes. In genetic terms, breeders aim to combine beneficial alleles into a single plant by crossing parental lines complementary for one or several traits of interest (to "mix"). The challenge is then to develop elite cultivars from "superior" plants (i.e. that outperform their parents for the desired traits) identified in the progeny. The traits of interest have then to be fixed in a genotype that performs well in field conditions ("to fix"). If this new genotype is promising enough, it can enter the evaluation process towards a potential commercial use as pure line or in hybrid combinations.

Plant breeders thus heavily rely on the transfer and reshuffling of genetic information that occur during meiosis. Meiosis is a specialized cell division that leads to the production of gametes. During meiosis, chromosomes are recombined through the formation of Crossing Over (COs), which are reciprocal exchanges of genetic information between homologous (maternal or paternal) chromosomes. Increasing knowledge has been gained on the underlying molecular mechanisms that govern meiotic recombination. A more comprehensive view has begun to emerge notably in plants with the contribution of model species such as *Arabidopis thaliana* (thale cress), and to a lesser extent *Oryza sativa* (rice) and *Zea mays* (maize). This has led to identification and functional analysis of more than 80 genes involved in meiosis (Mercier et al., 2015) and has paved the way for new strategies aiming at controlling CO formation in plants.

The intensity of genome reshuffling that occurs during meiotic recombination is a factor to be reckoned with in plant breeding. Because of its direct and indirect effect on genetic diversity, the intensity of meiotic recombination in a region influences how much diversity is available in the first place for breeders. The intensity of meiotic recombination also determines how much effort is needed to exploit the existing genetic diversity in plant breeding. Because breeding program objectives and strategies are quite diverse, there might be an interest to finely tune the level of meiotic recombination according to one's specific need. For example, the desired CO frequencies might not be the same depending on the source of the genetic variation that is used in the cross and the genetic architecture that govern the trait of interest (Figure 1).

For example, when dealing with the introgression of a single trait (Figure 1A - 1B), a high CO frequency will increase the odds of successfully introducing the desired trait into an elite genotype while reducing the size of the introduced fragment and thus the number of undesired alleles genetically linked to the desired traits. However, it will concurrently increase the number of non-desired regions that are introgressed genome wide from the source and thus possibly result in an increased effort to return to an elite genotype.

When working with genetically more complex traits (Figure 1C – 1D), high CO frequencies will increase the chance of finding new positive allele associations in the progeny of a cross but will also break some of the pre-existing positive associations found in the parents.

Although the duality of meiotic recombination (breaking up associations between both beneficial and detrimental alleles) is less likely to be problematic when the two parental lines are themselves fixed for agronomical trait of interest, genetic diversity is often introduced from exotic germplasm, which were not subjected to intense selection and are thus more likely to introduce undesired alleles in the progeny.

The control of CO patterning in plant is a burgeoning field. Identification of hyper-recombinants plants in *A. thaliana* suggests that recombination frequencies might become one of the many parameters a breeder can tweak to reach its objectives. However, there is still a huge leap that has to be made to translate these findings into crops and especially the allopolyploid ones. My work is one of the many steps towards that end.

In the first chapter of my manuscript, I will provide an overview of the meiotic process; I will notably give some insights into its molecular mechanisms and consequences on the evolution of plant genomes. In the second chapter, I will present my plant model, *Brassica napus*, and review some of the key aspects regarding its polyploid origins, its relevance for plant breeding and the control of meiotic recombination in this species. In the third chapter, I will introduce the objectives of my work. I will present my results in two chapters, in chapter four I characterize the main sources of variation I detected when performing a transcriptomic analysis on a single meiotic cell type. In chapter five, I present the outcome of a translational biology approach to produce hyper-recombinant plant in *Brassica*. I will then summarize and discuss these results in chapter six before giving the perspectives of my work.

Chapter 1: Bibliographic review

Within this chapter I will provide the necessary information that is needed to understand what will be presented and discussed in the manuscript. I want to stress that this chapter does not pretend to be exhaustive on any of the processes I will introduce. References are provided in the text to guide the reader to recent reviews that cover with more details some of the mechanisms presented here.

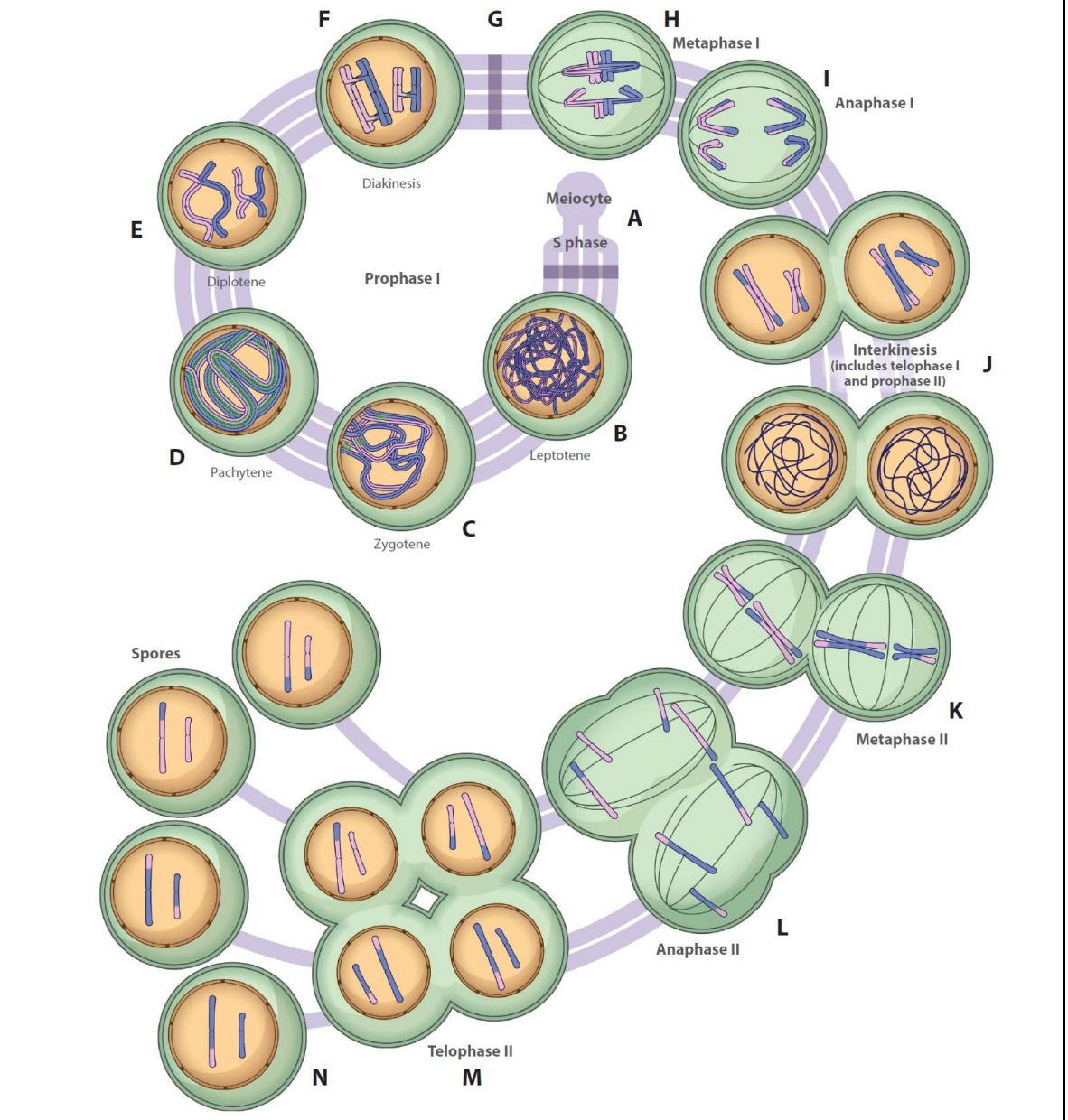
Below, I briefly outline several mechanisms, notably meiosis and meiotic recombination that constitute the biological background for my work. I notably present what are the general determinants that govern CO patterning and give further insight about how this control over CO formation is implemented in polyploids. I also describe how CO patterning determines how much of the genome diversity is available for crop improvement and review how gaining control on CO patterning might answer some of the challenges faced by breeders.

1.1 Progression of meiosis, as seen through the prism of chromosome association/segregation

Meiosis achieves segregation of maternal and paternal (i.e., homologous) chromosomes through two successive cellular divisions that are preceded by a single round of DNA replication (Figure 2). This is achieved through the bending of mitotic cell cycle rules to prevent an intervening replication between the two meiotic divisions (Wijnker and Schnittger, 2013). The first division (meiosis I; Figure 2B to 2J) allows separation of homologous chromosomes, while the second division (meiosis II; Figure 2J to 2N) leads to the separation of sister chromatids (i.e., the two identical copies of a single replicated chromosome).

Proper chromosomes segregation during meiosis is highly dependent on the establishment and removal of “connections” between sister chromatids and homologous chromosomes. The following is an outline of these “connections” as they occur during meiosis.

Cohesion between sister chromatids is established during DNA replication; it is mediated by a multi-protein cohesin complex that, according to the model proposed in Nasmyth and Haering, (2005), forms a ring structure within which sister chromatids are entrapped (Figure 3a and 3b). This structure holds sister chromatids together until their controlled separation at anaphase (see below).


F
G
H
Metaphase I
I
Anaphase I
Diakinesis
Meiocyte
A
S phase
E
Diplotene
Prophase I
Interkinesis
(includes telophase I
and prophase II)
J
B
Leptotene
D
Pachytene
C
Zygotene
Spores
K
Metaphase II
L
Anaphase II
Telophase II
N
M

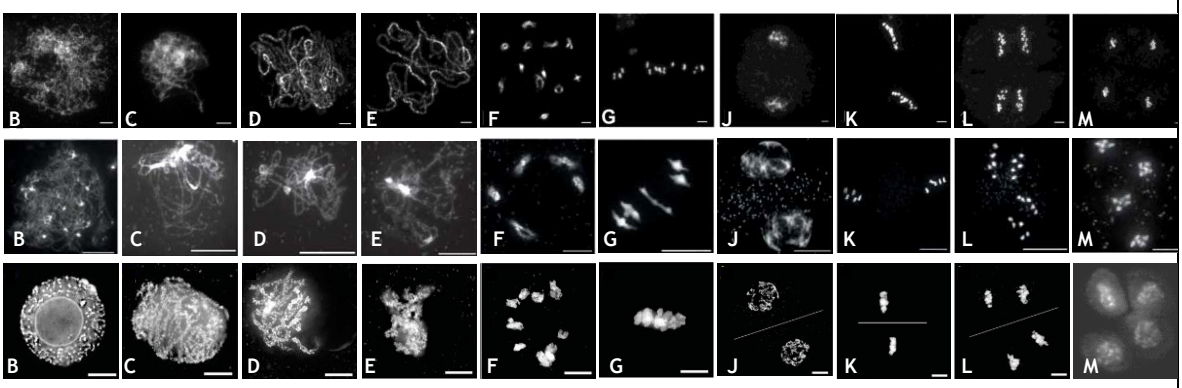

B
C
D
E
F
G
J
K
L
M

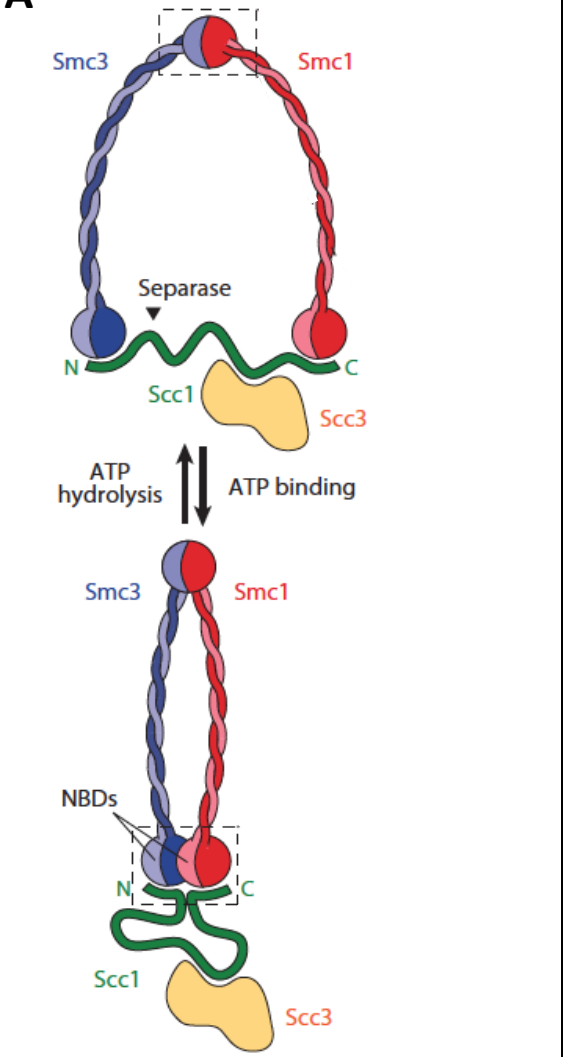
Figure 2: Progression of meiosis and cytological manifestation of meiosis stage in plant model species

[Adapted from (Hamant et al., 2006; Wang et al., 2009; Mercier et al., 2015)]

Chromosome structural changes during meiosis can be visualized in cytology with DAPI staining of male meiocytes, here in *Oryza sativa* (top), *Arabidopsis thaliana* (middle) and *Zea mays* (bottom).

Meiotic division is preceded by premeiosis, which encompasses meiocyte differentiation and meiotic S phase (the single round of DNA replication) (A). Each meiotic division comprises 4 stages (prophase, metaphase, anaphase, and telophase). The very first stage (prophase I) is the longest and is divided into 5 substages (leptotene, zygotene, pachytene, diplotene, and diakinesis). (B) During leptotene chromosome axes are formed and recombination is initiated. In cytology, chromosomes become visible as unpaired threads. (C) At zygotene, a proteinous structure (the synaptonemal complex (SC), see text) polymerizes between homologs bringing them in close apposition. (D) At pachytene, the SC is complete; all chromosomes are closely aligned with one another. (E) At diplotene, the SC disassembles. Crossovers connect homologous chromosomes. (F) At diakinesis, chromosomes are condensed; both sister chromatids and homologous chromosomes are connected to each other forming discrete and separate bivalents. Prophase I finishes and the nuclear envelope breaks down. (G) At metaphase I, all bivalents align on the metaphase plate. At anaphase I (I), the release of sister chromatid cohesion along chromosome arm allows migration of homologous chromosomes at opposing pole. Pericentromeric cohesion is specifically protected. (J) At interkinesis, two nuclei form and chromosomes briefly decondensate. This stage encompasses telophase I and prophase II. In monocotyledons, cytokinesis occurs before meiosis II starts; in dicotyledons, cytokinesis happens only at telophase II. (K) At metaphase II, two spindles form and align chromosomes on two metaphase plates. (L) At anaphase II, sister chromatids separate as a result of centromeric cohesion loss. (M) At telophase II, four nuclei form. (N) At cytokinesis, haploid spores are released.

Scale bar: maize and rice = 5µm, *A. thaliana* = 10µm.



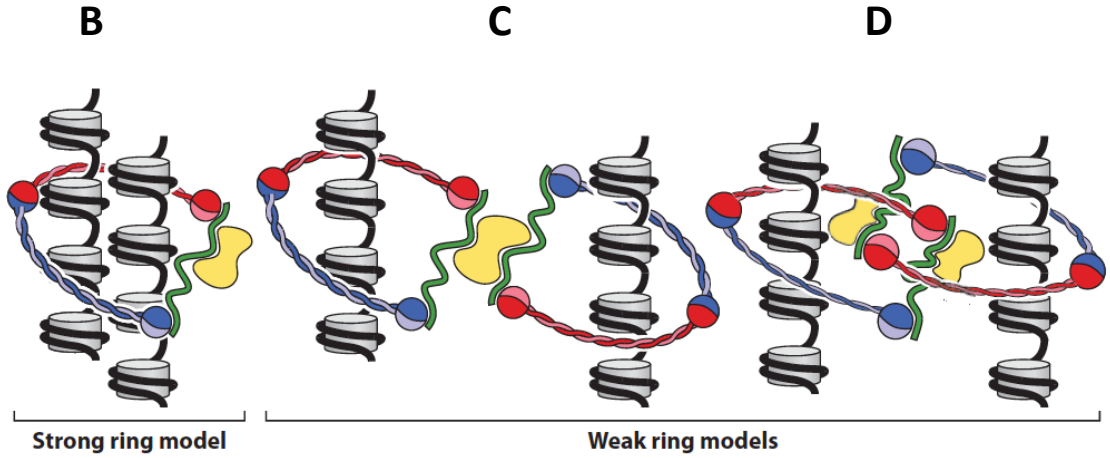


Figure 3: The cohesin complex [reproduced from (Nasmyth and Haering, 2009)]

(A) The cohesin complex is highly conserved in eukaryotes, it is composed of a core of four evolutionary conserved proteins, extensively studied in yeast and animals. In mitosis, the core cohesin complex consists of two SMC (structural maintenance of chromosomes) proteins, SMC1 and SMC3, and two auxiliary SCC (sister chromatid cohesion) subunits, SCC1 and SCC3. During meiosis, the structure of the cohesin complex is highly similar, except for the SSC1 component, which is replaced by its counterpart Rec8. In Arabidopsis, single copy homologs of SMC1, SMC3, SCC3 and Rec8 (named SYN1 but also DIF1 and AtRec8) have been identified in addition to 3 paralogs of SYN1 (SYN2, SYN3, and SYN4) whose exact roles have not been fully unrevealed yet (Zamariola et al., 2014). SMC1 and SMC3 consist of a globular head and a hinge domain, connected by a long anti-parallel coiled coil. Heterotypic interactions between the hinge domains of SMC1 and SMC3 lead to the formation of V-shaped SMC1/3 heterodimers with an SMC1 nucleotide-binding domains (NBD) at the end of one arm and an SMC3 NBD at the end of the other.

(B-D) The discovery that Smc1, Smc3, and Scc1 form a ring has led to the formulation of the ring model. Two versions have been proposed. (B) In the strong ring model, sister chromatids are trapped inside a single monomeric cohesin ring. In the weak ring models, sister chromatids are either (C) held together by interactions between two different rings, one that has trapped one chromatid and a second that has trapped the second or (D) held together by interactions with interconnected cohesion rings

The latter is attributable to the fact that meiotic cohesins are prominent components of the meiotic chromosome axis, the integrity of which is compulsory for meiotic recombination (Storlazzi et al., 2008). In fact, from the outset of leptotene, sister chromatids form linear arrays of loops (Figure 4), the bases of which comprise a structural axis delineated by the "axial element" (AE) (Zickler and Kleckner, 2015). This axis comprises a complex meshwork of protein/DNA interaction including cohesins, condensins and specific AE proteins such as ASY1 in *Arabidopsis thaliana* (Armstrong et al., 2002).

Chromosome interactions between homologues begin at leptotene, when chromosomes start searching for a partner to align and recombine with (Zickler and Kleckner, 2015). The early steps of homologue recognition and alignment are still not clearly understood but they probably involve numerous processes including: dynamic chromosome movement, clustering of telomere to the nuclear envelope (the so-called bouquet) and the early steps of meiotic recombination [reviewed in (Zickler and Kleckner, 2016)]. Indeed, nascent recombination intermediates ensure periodic inter-homologs local contact through the formation of "bridges" between chromosomes during their search for homology. Thus, recombination in most organisms plays a central mechanistic role by contributing to homologue recognition and by bringing homologs together in space, i.e., homolog pairing.

While recombination progresses during zygotene, the homologous chromosomes start to 'zip up' (i.e. synapse), as a proteinaceous structure, the synaptonemal complex (SC), forms between them (Figure 4). The SC has long been recognized as a hallmark cytological feature of meiosis. The SC is a tripartite structure that comprises the AEs of the two homologous chromosomes, which are now called lateral elements (LEs), and a central element (CE) that consists of transverse filaments interconnecting the two lateral elements.

The central-element proteins are poorly conserved at the sequence level. However, the transverse filaments of the SC central region display in all organisms the canonical structure of the large coiled-coil protein Zip1 in yeast. Although several other components of the CE have been identified [reviewed in (Zickler and Kleckner, 2015)], it is not known how they interact to form the SC. The function(s) of the SC itself is/are not completely understood and it is believed that the SC has both global roles in maintaining chromosome order within the nucleus and local roles at sites of recombination (Zickler and Kleckner, 2015).

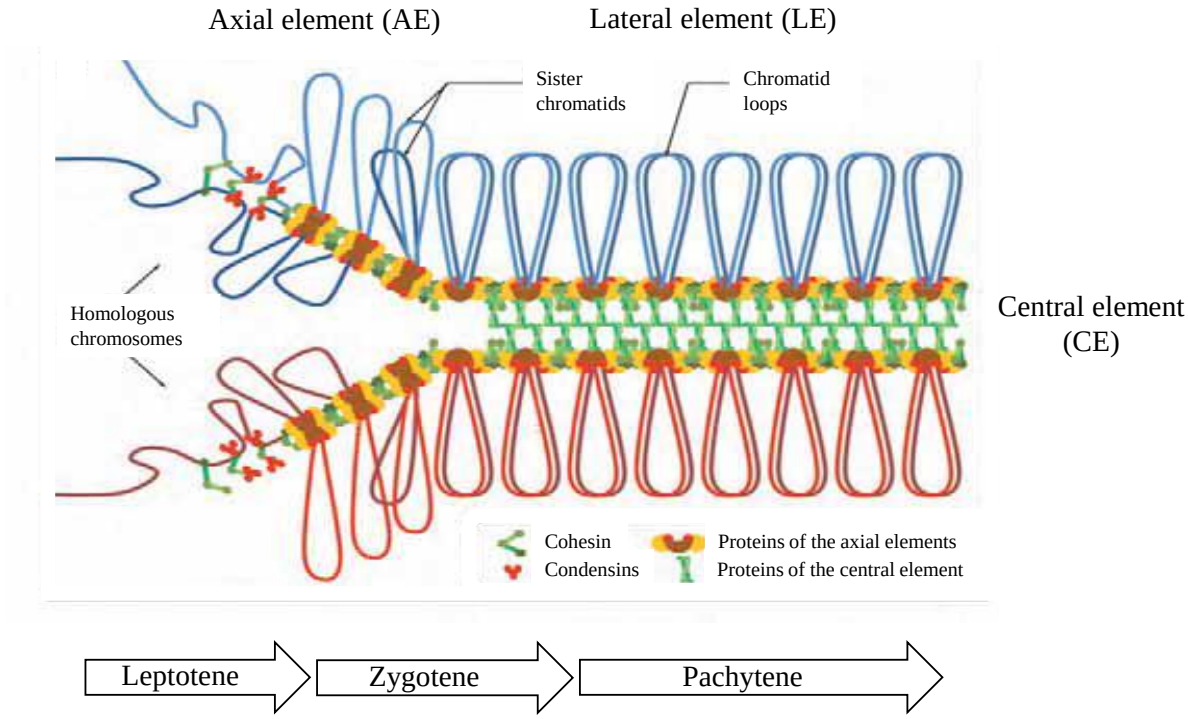


Figure 4: The Synaptonemal complex (adapted from Macaisne 2010; Page and Hawley, 2003)).

During meiotic prophase, chromosomes are organized into linear arrays of chromatin loops, the bases of which define chromosome axes (composed of cohesins, condensins and other proteins). The period of time when the Synaptonemal Complex (SC) is forming defines zygotene. SC initiates non-randomly and at multiple location along the chromosome in plants. The nucleation of the SC occurs at site of inter-homologs local contact mediated through pairing. SC is composed of Lateral Elements (LEs) that are formed from the formerly named Axial elements (AE)s and a central element (CE). One of the components of the CE is the transverse filament that comprised elongated protein dimers interacting with both of the LEs and with each other. The presence of a complete SC defines pachytene.

The SC is completed by pachytene, at which time homologous chromosomes are fully synapsed. At this stage, at least in yeast, meiotic recombination is completed (i.e. before the end of pachytene); it generates two distinct types of products: the crossovers (COs) and the non-crossovers (NCOs). Meiotic COs consist of reciprocal exchanges of genetic material over large chromosome intervals while NCOs involve only a unidirectional transfer of genetic information over short intervals (see below).

At the onset of diplotene, the SC breaks-down and releases homologous chromosomes except at the sites where COs have occurred. These physical connections, which are the manifestations of COs and sister chromatid cohesion, are known cytologically as chiasmata. The pairs of homologous chromosomes, each made of the two-replicated sister chromatids, are thus physically linked together and form a structure unique to meiosis, called a bivalent.

At the end of prophase I, bivalents are maximally condensed and align on the equator of the metaphase I plate. Sister chromatid cohesion is then lost in a stepwise process. At anaphase I, sister chromatid cohesion is released from chromosome arms but preserved in centromeric regions. This allows the recombinant homologous chromosomes to migrate to opposite poles. The active protection of cohesion between sister chromatids allow their proper segregation during a second “mitotic-like” division at which point centromeric cohesion is lost, allowing the sister chromatids to separate to form a tetrad of four haploid spores (Figure 2).

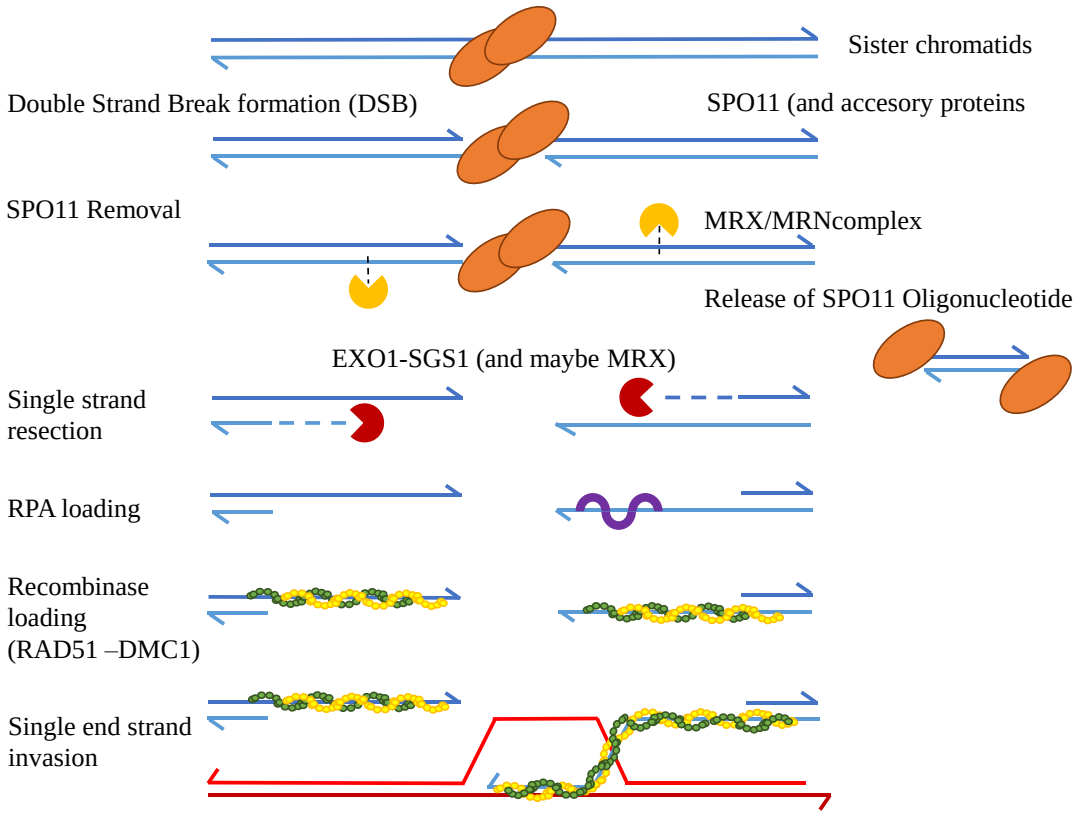


Figure 5: Early Double Strand Break (DBS) Repair Process repair process in budding yeast (adapted from (Neale et al., 2005; MacQueen, 2015).

SPO11 and its accessory proteins are recruited to the chromatin and create a DSB. SP011 remains associated with chromatin until breaks driven by MRX/MRN complex free the SPO11-nucleoprotein filament on either side of the DSBs. The DNA is then resected by EXO1-SGS1 (and maybe MRX) from 5' to 3', which frees the 3'end of the complementary strand. RPA is then bound and load the recombinases RAD51 (yellow) and DMC1 (green). It is not known whether only one (as represented here) or both 3′ ends at a DSB exhibit equivalent homology search behavior (Brown et al., 2015).

1.2 The molecular mechanisms of meiotic recombination

Meiotic recombination is initiated during leptotene through the formation of programmed double strand breaks (DSBs). The core mechanism for DSB formation is shared in eukaryotes. DSB formation is catalysed by SPO11, an evolutionary conserved protein, whose catalytic complex shares similarity with the archaeal topoisomerase VI (topo VI). In *Arabidopsis* two non-redundant SPO11 homologs (SPO11-1 and SPO11- 2) are similar to the A subunit of topo VI. A newly discovered homolog for the B subunit of topo VI mediates the interaction between the two SPO11s (Vrielynck et al., 2016). SPO11 is not sufficient for DSB formation but requires a set of essential partners that are poorly conserved between species [reviewed in (Lam and Keeney, 2015)].

Once formed, DSBs have to be repaired in an error free manner. Most of what we known about the mechanisms by which DSBs are repaired has been demonstrated in *Saccharomyces cerevisiae* (budding yeast) only. Unless specified otherwise, I will refer to these data.

DSBs repair is initiated by the nucleolytic processing of the 5′ ends of DSB through a multi-step process called DNA end resection. First, SPO11 which remains covalently attached to the 5’ -ends of the DNA on either side of the break site is removed. Removal of SPO11 by endonucleolytic cleavage is carried out by the MRX/MRN complex (MRE11/RAD50/XRS2 or MRE11/RAD50/NBS2) together with COM1/SAE2. This releases short oligonucleotides bound to SPO11 (Neale et al., 2005). DNA is then further resected to generate 3′ single stranded DNA overhangs which are subsequently bound by RPA (Replication protein A) and loaded by the recombinases RAD51 and DMC1. The resulting nucleoprotein filaments invade duplex DNA to carry out homology searches and form heteroduplex to initiate strand exchange (Figure 5).

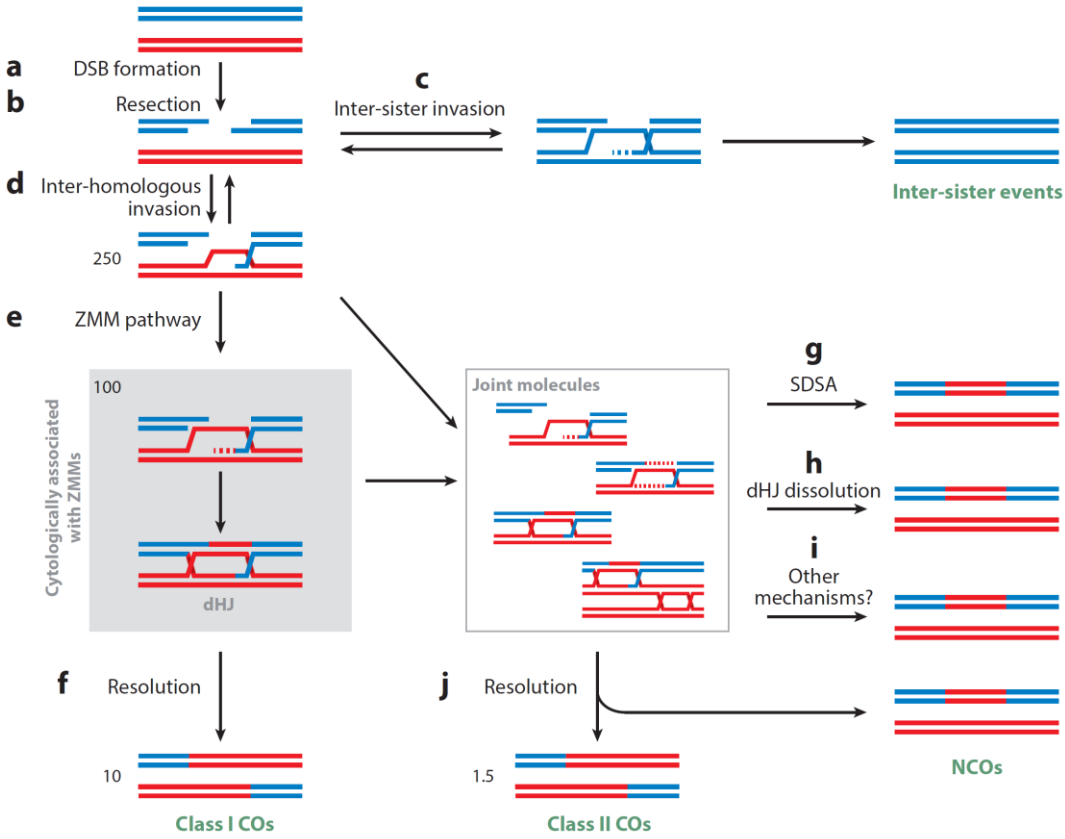


Figure 6 : Meiotic recombination mechanisms [reproduced from (Mercier et al., 2015)]

Meiotic recombination is initiated by a large number of double strand breaks (DBSs) that are processed to yield single 3'-OH single-stranded DNA. DSBs can be repaired using either sister chromatid as a template (c) or one of the two homologous chromatids, forming a D-loop (d). Most of the recombination intermediates are turned into non cross-overs (NCOs) through distinct mechanisms (g, h, i). Alternatively, recombination intermediates can yield class I or class II COs when they are taken in charge either by the ZMM (f) or the MUS81 depending pathway (j)

RAD51 is essential for DSB repair in mitotic recombination (using the sister chromatid as a repair template) but functions as a DMC1 accessory factor during meiotic CO formation. DMC1, which is only active during meiosis, is thus mediating the main pathway for DNA repair [reviewed in (Mercier et al., 2015)] and promotes inter-homologue recombination. There are indeed three possible templates for DSB repair: the sister chromatid and the two non-sister chromatids from the homologous chromosome. During meiosis, homologous strands are preferentially used as a template; this process, known as the inter-homologous bias, is mediated by a series of proteins, including DMC1, ASY1 and a few others [reviewed in (Mercier et al., 2015)]. Inter-homologous bias is instrumental (and required) to the formation of at least one CO per homologous pair, i.e. the obligate CO.

Invasion of an intact chromatid by 3′ single stranded DNA forms a displacement loop (D-Loop), which is extended by DNA synthesis. The resulting heterologous duplexes are likely very unstable; they may thus dissociate after a short elongation (invasion/dissociation can even go back and forth several times; (Symington and Heyer, 2006)) and then be repaired by synthesis-dependent strand annealing (SDSA). This pathway is considered as an important route for Non Crossover (NCO) formation (Allers and Lichten, 2001) (Figure 6).

Some nascent inter-homologous intermediates can also get stabilized by components of the ZMM pathway, which involves a group of proteins first described in yeast (Zip1, Zip2, Zip3, Zip4, Msh4, Msh5 and Mer3). This allows the capture of the second end to generate a double Holliday junction (dHJ), a cross-strand recombination intermediate that can be resolved into class I crossovers (COs). Even if numerous early recombination intermediates are at first processed by ZMMs, only a few mature into COs (Figure 6). These COs are interference-sensitive; this means that they tend to localize farther apart along the chromosome than expected by chance. They account for the majority of CO in plants (Mercier et al., 2005). Finally, it is worth pointing that two other conserved proteins, MLH1 and MLH3, act in the ZMM pathway although they are not classified as ZMMs. This will be useful in the next section of the document.

Alternatively, the joint molecules can be processed through dissolution, which results in a NCO, or acted upon by enzymes, such as MUS81, to form interference-insensitive COs (Figure 6). These class II COs account for 10% of overall COs in *A. thaliana* (Higgins et al., 2008).

In addition to all these pro-CO activities, a set of proteins showing anti-CO activities has been recently identified. These proteins channel the vast majority of joints molecules towards NCO (Sanchez-Moran et al., 2007); they can be classified into three distinct pathways, all of which limits class II CO formation (Figure 6).

(i) The helicase Fanconi Anemia Complementation Group M (FANCM) is thought to promote NCO formation through the SDSA pathway (Crismani et al., 2012). Although FANCM acts as a landing pad for multiple Fanconi Anemia (FA) associated proteins, only its direct DNA-binding cofactors MHF1 and MHF2 limit CO formation at meiosis (Girard et al., 2014). FANCM utilizes its DNA-dependent ATPase activity to translocate along DNA and promote the migration of Holliday junctions. Mutants defective in ATPase activity are unable to process the Holliday junction and are similarly defective in D-loop dissociation (Gari et al., 2008).

(ii) The topoisomerase3α (TOP3α) and the RECQ4 helicase promote NCO formation via D-loop displacement and SDSA independently of FANCM, possibly by unwinding different JM substrates (e.g., nascent versus extended D-loop) (Séguéla-Arnaud et al., 2015).

(iii) The AAA-ATPase (ATPases Associated with diverse cellular Activities) FIGL1 is thought to act earlier during invasion step. It has been proposed that FIGL1 prevents the formation of aberrant joint molecules by regulating strand invasion (Girard et al., 2015).

Thus the number of COs can be viewed as the result of pro- and anti-CO activities, which are mediated by main actors of the recombination machinery.

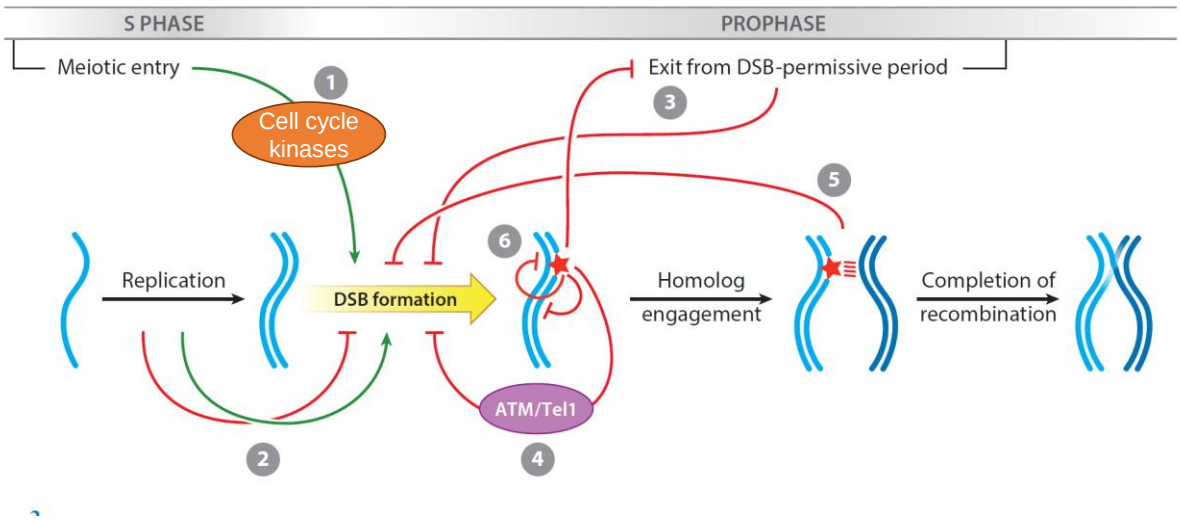


Figure 7: Circuits that regulate double strand breaks timing and patterning [adapted from (Keeney et al., 2014)]

Multiple regulatory feedback loops exist that limits DSB formation to a permissive period. DSB formation is promoted (green arrow) by actors of the cell cycle (1) at the onset of meiosis and starts after a fixed time period following DNA replication (2). In face of replication problems, DSB machinery can be downregulated (red arrow). As meiosis progresses though prophase stages, the window of opportunity for DSB formation closes as recombination intermediates form (3). DSB formation itself activates a retro control loops through the DNA damage sensitive ATM (ataxia telangiectasia mutated) kinase (4). The formation of higher meiosis specific structure acts as a signal to inhibit further DSB formation (5). Interference in cis (along the same DNA molecule and in trans between sister chromatids or homologous chromosome influence DSBs distribution (6).

1.3 The progression of meiosis is intertwined with meiotic recombination

For sake of clarity, I have so far presented the progression of meiosis and the molecular mechanisms of meiotic recombination separately. However, these two processes are interdependent in many organisms. This is notably illustrated by the interplay between meiosis progression and DSB formation.

Temporal control is imposed over DSB formation to ensure that they will occur after DNA replication (when sister chromatids exist) and will stop before first chromosome segregation. At the onset of meiosis, DSB formation is directly promoted by key drivers of the cell cycle and coordinated with DNA replication (Figure 7). In *S. cerevisiae,* DSB formation occurs after a fixed period of time that follows replication, any delay in replication resulting in delayed DSB formation (Borde et al., 2000). This opens a window of opportunity for DSB formation, which is necessary for chromosome associations. As meiotic recombination progresses, this window progressively closes as retro-control loops activate. Multiple levels of regulation integrate meiosis progression and the duration of the DSB permissive state within the cell (Figure 7).

The formation of recombination intermediates acts as signal to reduce the formation of further DSBs. This was shown notably in *Caenorhabditis elegans* where the DSB permissive state is extended when the distribution or the number of CO intermediates is defective This echoes observation in *A. thaliana* where mutants defective for SC and class I CO show an increase in DMC1 foci (Chelysheva et al., 2007), which may suggest a prolonged DSB phase .

The SC in most organisms depends on DSB formation and in return regulates further DSB formation. For example completion of synapsis acts as a signal to stop DSB formation in mice (Kauppi et al., 2013). It has been hypothesised that formation of the SC renders chromosomes unfit substrates for SPO11. This could be mediated through the displacement of DSB-promoting factors like HORMA-domain proteins (like ASY1 in *A. thaliana*) from the axes soon after synapsis finishes. The exact role played by the different component of the SC (both AE and CE) with regards to DSB formation and maturation into CO has not been fully elucidated yet.

- In *A.thaliana*, mutants lacking the AE protein ASY1 show normal level of DSBs but ASY1 and ASY3 are required to promote maturation of recombination intermediates into crossover products (Armstrong et al., 2002; Sanchez-Moran et al., 2007; Wang et al., 2010a). In maize, DSY2, an ortholog of Arabidopsis ASY3, plays a double role being essential both for normal levels of DSBs and SC formation.

- In *C. elegans*, partial depletion of SYP-1, one of the few known SC component in this organism, alters CO distribution. SYP-1 acts by promoting local CO formation (the presence of SYP-1 being necessary at the site of a recombination intermediate to achieve a crossover), but also by inhibiting the formation of multiple COs per chromosome. The increased numbers of double COs observed in *syp-1* mutants could result from increased DSB formation in response to incomplete synapsis (Hayashi et al., 2010). The role of CE proteins is however not conserved across species; this is exemplified in plants with ZYP1. Whereas in rice, ZYP1 limits CO formation (Wang et al., 2010b), it has been shown to have the opposite role in barley (Barakate et al., 2014) while it could prevent non-homologous recombination in *A. thaliana* (Higgins et al., 2005).

These mechanisms that link DSB formation to the progression of meiosis come in addition to other regulatory mechanisms that regulate DSB formation. DSB formation itself is subject to retro-control loops as suggested by studies in mouse, flies, and yeast (Lange et al., 2011; Joyce et al., 2011; Zhang et al., 2011). This feedback mechanism is mediated by activation of the kinase ATM (ataxia telangiectasia mutated) upon DSB formation. The underlying mechanisms are still not well understood. In budding yeast, the retro-control loop acts both in cis, the occurrence of a DSB suppressing adjacent DSB formation over domains that span 100kb (Garcia et al., 2015), and in trans; for a given initiation only one DSB site is formed per four chromatids in a tetrad while one per pair is formed in *atm* mutants (Zhang et al., 2011).

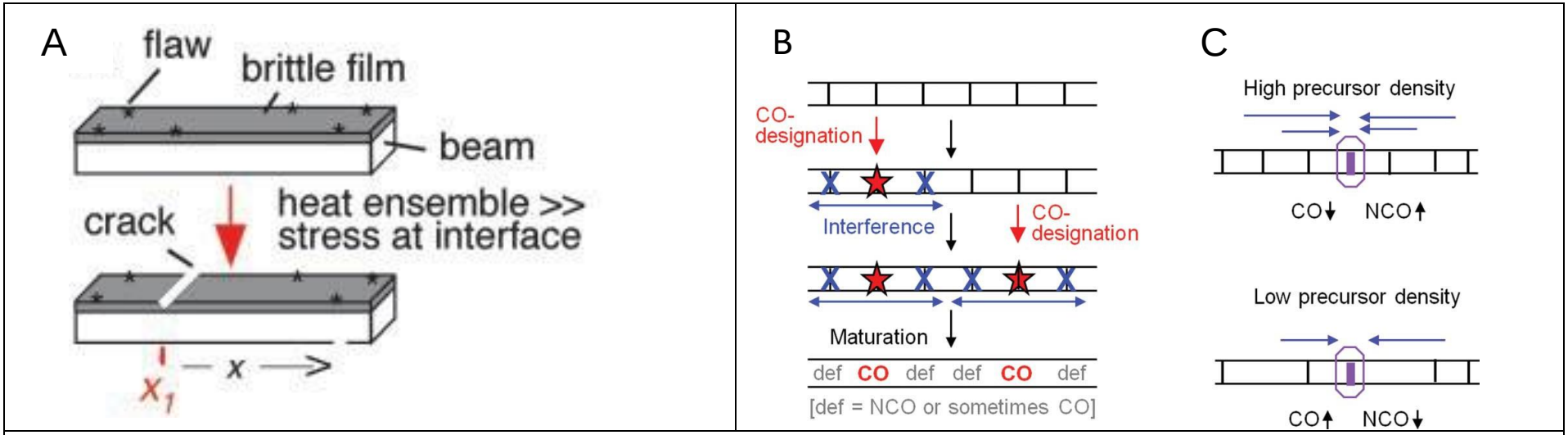


Figure 8: The beam-film model [reproduced from (Kleckner et al., 2004; Zhang et al., 2014)]

(A) Underlying physical model of the beam film model: An elastic beam or plate of metal is coated on one face by a thin brittle film of ceramic that contains flaws (black star) along its edges. If this ensemble is heated, the metal plate will have a greater tendency to expand than the overlaid ceramic film. If the two entities are tightly bonded, expansion of the metal plate will force the film to stretch. Heating gives rise to high tensile stress in the film that can trigger crack nucleation at the edge flaws (noted X_1). Once triggered, a crack extending down to the film interface propagates across the entire width of the ensemble from one edge to the other (black arrow). **(B) CO designation under the logic of the beam film model**: A chromosome with an array of precursors (vertical black lines) come under mechanical stress along its length. Eventually a stress promoted molecular change designates a first precursor to mature as CO (red star). This result in the propagation of a stress release signal (interference) that dissipates with distance. Additional precursors can then be matured as COs in regions where stress remains high. **(C) Crossover homeostasis from the perspective of an individual DSB-mediated precursor:** At high (low) precursor (vertical black lines) density, a precursor will be more (less) affected by the spreading interference signal (blue arrows) from nearby crossover-designations and thus will be less (more) likely to become a crossover.

1.4 The patterning of meiotic COs formation

It has been long observed that CO localization and number are tightly controlled. This has led to formulate three main features that rule CO patterning. First, because CO are needed to ensure proper chromosome segregation, every chromosome pair acquires at least one CO, hence called the "obligatory" CO. This occurs irrespective of chromosome length. Second, COs tend to be evenly spaced along chromosomes, the presence of one CO at a given position reducing the probability to observe a second CO in the vicinity. This phenomenon is being referred to as "CO interference". Third, "CO homeostasis" buffers the system against deficits (or excesses) of DSBs or precursor interactions, the number of COs being decreased (or increased) less than proportionally.

It has been hypothesized that these three features are different manifestations of a single patterning process that involves accumulation, local relief and redistribution of mechanical stress (Wang et al., 2015). In that model, called the ‘‘beam-film’’ model by analogy with a known physical system that exhibits analogous behaviour (Figure 8), chromosomes are under mechanical stress along their length as chromatin alternates between expanded and contracted states during meiosis (Kleckner et al., 2004). Occurrence of a CO then results in a local stress relief that tends to propagate in a manner that decreases gradually with distance. This disfavours maturation of additional COs in the regions where the stress level has been reduced. In other words, commitment of a precursor to become a CO leads to propagation of a signal that inhibits maturation of nearby CO. In that model, the "obligatory" CO is a consequence of accumulating stress that has to be relieved and CO homoeostasis is a reflection of interference strength (Zhang et al., 2014) (Figure 8).

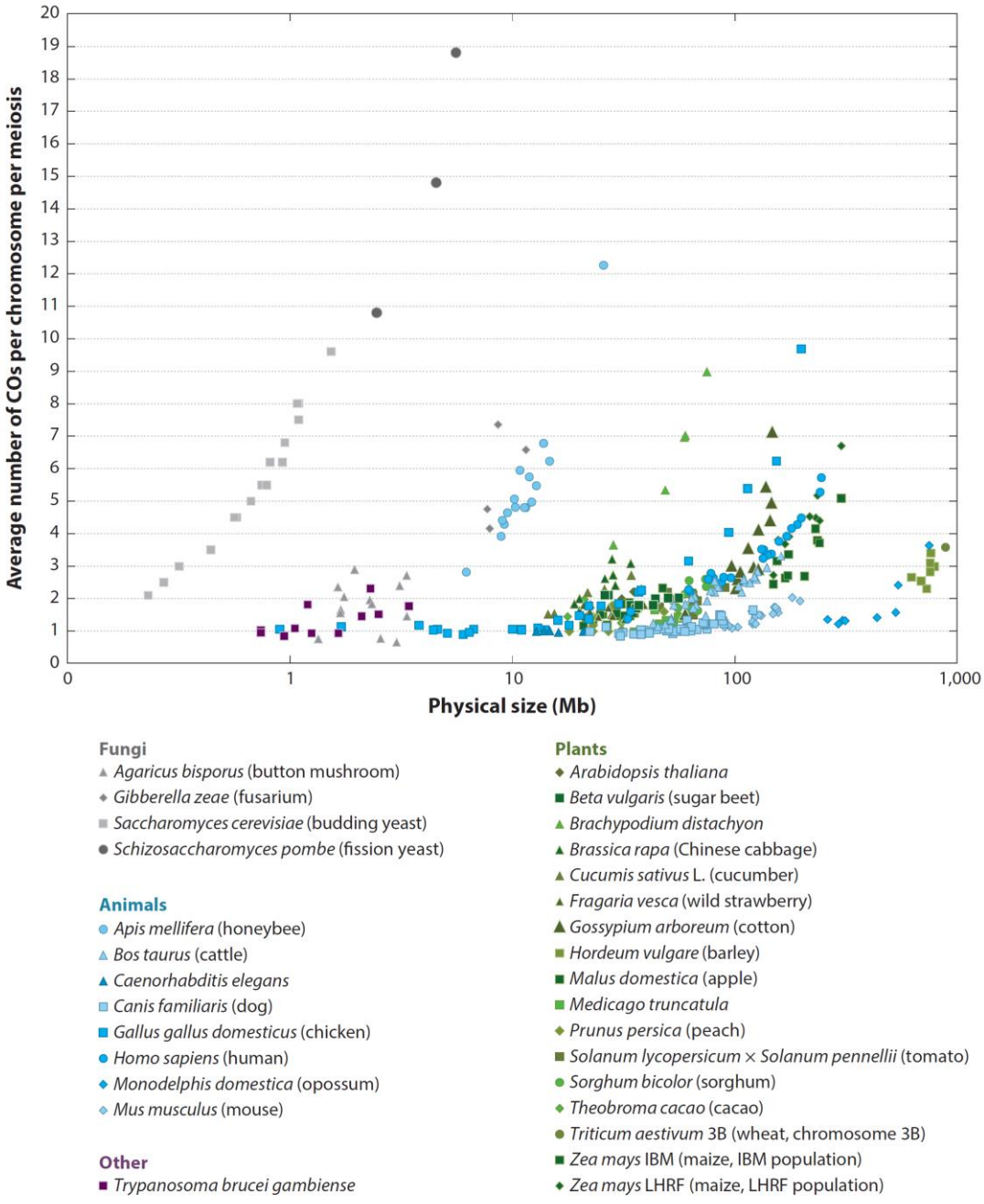


Figure 9: Number of crossovers (COs) per chromosome per meiosis in a variety of eukaryotes [reproduced from (Mercier et al., 2015)]

The number of COs, deduced from male/female-average genetic maps, is plotted against the physical size of each non sex chromosome (Mb, log scale). Irrespective of chromosome size, most of chromosome display less than 3 COs.

These rules of patterning combined with the existence of various pathways that actively limit CO formation (see above) result in a low average number of COs per chromosome; in the vast majority of species, the mean number of COs per chromosome is always above the obligatory CO but rarely exceeds three CO per bivalent. This holds true irrespective of the physical size of the chromosome and despite excess in CO precursors (Figure 9).

Because of interference, the first "obligate CO" that occurs will shape overall CO distribution. CO localization is quite variable from cell to cell but CO distribution along chromosomes is not homogeneous. Domains with high CO rates (hot regions) alternate with domains where CO rates are significantly lower than genome-wide average (cold regions). Although variation in DSB distribution must influence the observed heterogeneity in CO localization, only a partial correlation is found between DSB and CO frequencies in mice and humans (Smagulova et al., 2011; Pratto et al., 2014). Indeed, only a fraction of DSBs are repaired as COs and the factors that influence DSB fate are major contributors to the CO landscape.

Chromosomal primary structure is a great constraint to CO localization. In plants such as tomato, wheat, maize and barley, the large heterochromatic pericentromeric regions are almost completely devoid of COs, which are being restricted to distal euchromatic regions [reviewed in Mézard et al., 2015; Rodgers-Melnick et al., 2015]. In maize, the ratio of COs to DSBs strongly increases (5-fold) from centromeric to telomeric regions, contributing to the higher frequency of COs in distal regions compared to proximal regions (Stack and Anderson, 2002). In barley, CO formation correlates on where and when recombination is initiated. Higgins et al., 2012 observed that recombination was initiated throughout the entire nucleus, although in a polarized way. Recombination initiation in proximal regions occurs later than in the most distal ones and rarely progressed to yield COs reflecting a pronounced temporal differentiation in CO initiation and progression through meiotic prophase. As a result, ~half chromosome arms do not form COs in barley (Higgins et al., 2014). In other species, the regions where CO frequencies are low represent as much as 62% of the genome of maize (Rodgers-Melnick et al., 2015) and this proportion is even more extreme in wheat where 87% of the chromosome 3B is deprived in CO (Choulet et al., 2014). However, this observation is not universal. In *Allium fistulosum* for example, recombination is the highest in proximal regions and the recombination-rate gradient along chromosomes is reversed (90% of CO occurring within the proximal 25% of the SC length (Albini and Jones, 1987).

The relationship between recombination frequencies in a given region and the relative position of this interval along the telomere-centromere axis has been further studied in the grass family. In wheat (Lukaszewski et al., 2012) and wheat-rye hybrid (Lukaszewski, 2008), the authors observed recombination along chromosomes with an inverted arm, where the originally distal and CO-prone region has moved close to the centromere. As a result of this inversion, the pattern of CO distribution was also inverted, with recombination being higher in the region that is positioned next to the centromere. In another study in wheat, (Jones et al., 2002) brought the position of proximal CO-poor regions closer to the telomere by deleting the most distal part of a chromosome. This resulted in an increased recombination frequency in this newly defined terminal segment as compared to CO rate in the same segment of the complete arm. Altogether, this suggests that genomic composition is a main but not the sole determinant for the ability of a chromosome region to recombine.

There are other factors than chromosome primary structure that contribute to shape CO landscape. This is best illustrated when comparing CO pattern between male and female meiosis. Whereas in tomato CO number and distribution along chromosomes do not depend on sex, this is not true in most cases (Lenormand and Dutheil, 2005). The pattern can be very contrasted as in *A. thaliana* for example where CO rates in distal regions are very high in male meiosis but very low in female meiosis (Giraut et al., 2011).

Relatively recently, chromosome secondary structure has been shown to impact CO localization. Several lines of evidence highlight the impact of epigenetic marks on DSB formation. The general idea is that DSBs tend to clusters in region where DNA is accessible. This preference is not driven by SPO11 (the main catalyser of DSB formation) itself as it displays no or little DNA sequence specificity (Prieler et al., 2005). While in mammals SPO11 is guided via the histone-trimethyltransferase PRDM9 to consensus sequence (Baudat et al., 2010), this is not the case in plants, which lack PRDM9. A series of marks that can differ between species were instead found to correlate with DSB rich sites (Coopera et al., 2016). In *Arabidopsis* and maize, for example, genome-wide analysis of hotspots show low levels of DNA methylation (Choi et al., 2013; Rodgers-Melnick et al., 2015).

In addition to chromosome primary and secondary structure, recombination frequencies also seem to depend on sequence identity. CO frequencies tend to decrease between regions displaying increased sequence divergence. For example reduced recombination frequencies are observed in regions where introgressions from related species are present in heterozygous state (for example in tomato see Liharska et al., 1996; Canady et al., 2006). Reduction in recombination frequencies is more pronounced when the introduced fragment is from a species that is more distantly related to the recipient. This general rule suffers exception as observed in *A. thaliana* in a context of lesser sequence divergence (crosses between *A. thaliana* accessions). Ziolkowski et al., 2015 observed an increase in recombination frequencies within heterozygous regions (and a decrease in homozygous regions) in situation where homozygous and heterozygous regions were juxtaposed. The occurrence of natural structural variation may also supress local recombination in maize as hypothesized in (Bauer et al., 2013 and Rodgers-Melnick et al., 2015.

The reduction in CO frequencies that is observed in context of sequence divergence is reminiscent of what is observed in polyploid species where CO preferentially form between the most closely related genomes.

1.5 How to deal with the polyploidy situation?

While the control of meiosis in a diploid cell is already an intricate process, it becomes even more difficult in a polyploid cell where every chromosome has more than one possible partner to recombine with. The presence of more than one (closely) related genome leads to unbalanced chromosome segregation, aneuploid gametes and reduced fertility whenever illegitimate or multiple recombination occurs (Ramsey and Schemske, 2002).

Polyploids fall into two broad categories according to their mode of origins. Autopolyploids are formed by the doubling of a single diploid genome within a species while allopolyploids have a hybrid origin. In allopolyploids, pairs of homologous chromosomes coexist with more diverged chromosomes that originated by speciation and were brought back together in the same genome (homoeologs) (Glover et al., 2016). In numerous polyploids, there is no complete preference of homologous over homeologous recombination, the mutual affinities depending on the relative relatedness of the genomes involved (Wu et al., 2001). This is further exemplified in modern sugarcane (*Saccharum* spp.), a complex polyploid displaying unsystematic meiotic behaviour (Jannoo et al., 2004). Observation of meiosis in polyploids species has given insights into the mechanism that control meiotic recombination in order to achieve a balanced, stable meiotic division.

For autopolyploids, all copies being the same, there is no basis for preferential chromosome recognition. Proper chromosome segregation then relies on the random assortment of homologs into pairs instead of multivalents. Although exceptions exist (reviewed in Bomblies et al., 2016), reduction in multivalent formation occurs generally through a reduction in the overall CO frequency. Accordingly, it has been observed that established autopolyploids make less COs than newly formed autopolyploids (Yant et al., 2013) and numerous autopolyploids tend to do no more than the obligatory CO (Bomblies et al., 2016 and references within). It has been proposed that this could be the consequence of a strong interference (over a distance comparable to chromosome length) that would ensure that no chromosome becomes connected to more than one partner and that each chromosome will display at least one CO (Bomblies et al., 2016). More insight over the molecular basis of this control has been gained recently in the autotetraploid *Arabidopsis arenosa*. Yant et al., 2013 used a genome scanning approach to compare the genome of diploid and tetraploid *A.arenosa* and detected evidence of selection for 39 regions spanning 44 genes of which 8 were meiotic genes.

These genes encode the chromosome axis components ASY1 and ASY3, the cohesins and cohesin-associated proteins SMC3, SYN1 (Rec8) and PDS5 as well as the synaptonemal complex transverse filament proteins ZYP1a and ZYP1b). Ongoing work aims at determining whether the alleles that have been selected for in the autotetraploids tend to decrease CO frequencies.

In allopolyploids, proper chromosome segregation not only requires that chromosomes form pairs instead of multivalents but also that the pairs are restricted to homologous chromosomes. The underlying molecular mechanisms that inhibit CO formation between homoeologs are not well understood; only the *Ph1* locus in allohexaploid wheat (AABBDD) has been characterized at the molecular level so far. *Ph1* corresponds to a cluster of cyclin dependant like kinases (CDKs) on chromosome 5B. Although 5B CDK-like genes are transcribed, they all seem to be defective (Greer et al., 2012). Deletion of the Ph1 region from chromosome 5B results in increased expression of the corresponding CDKs on the homoeologous chromosome 5A and 5D (Al-Kaff et al., 2008). This has led to the assumption that *Ph1* could reduce the overall Cdk activity and that this reduction in activity would result in CO suppression between homoeologues. In agreement with an increase of Cdk2-type activity in the absence of *Ph1*, treatment with okadaic acid, a drug that increases Cdk activity, was shown to increase CO formation between homoeologous chromosomes (even in the presence of Ph1) thereby phenocopying the effect of deleting Ph1 (Knight et al., 2010). Likewise, Greer et al., 2012 found an increased level of the histone H1 phosphorylation, one of the best-characterized Cdk2 substrates, in a mutant defective for Ph1.

Martín et al., 2014 assessed whether the dynamics of synapsis and the loading of the recombination machinery (though immunolocalization of MLH1 foci) was affected by *Ph1*. In the presence of *Ph1,* the number of MLH1 sites correlates with the number of chiasmata, the cytological manifestation of a CO. When *Ph1* is absent, and although some chromosome arms lack chiasmata, no decrease in the number of MLH1 sites was observed. In wheat-rye hybrids, where only homoeologous recombination can occur, Martin et al. (2014) found no correlation between chiasmata number and MLH1 sites both in the presence and in the absence of *Ph1*. They concluded that *Ph1* acts by preventing MLH1 sites on synapsed homoeologues from becoming COs later in meiosis.

In another study, Boden et al., 2009 observed a phenotype reminiscent of *ph1* in transgenic lines showing a reduction in the level of *TaASY1* transcription. *TaASY1* is the wheat homologue of *ASY1 (*an axial element protein) in *A. thaliana*. More interestingly, the authors also provided evidence that *ASY1* is strongly up regulated (20 fold) in the absence of *Ph1* during pre-meiotic interphase and leptotene to pachytene. Although, only a correlation has been drawn so far, it is tempting to imagine that the activity of *Ph1* could be mediated by the fine tuning of the expression of a network of meiotic genes.

Unlike *Ph1,* the locus *Ph2* on the chromosome 3D of wheat is not directly involved in the suppression of CO between homoeologs (Martinez et al., 2001). *Ph2* mutants are delayed in the progression of synapsis, which would result in an incomplete action of the *Ph1* locus to prevent homoeologous recombination. The identity of *Ph2* remains elusive, the only information available so far is a list of 218 genes putatively falling in the Ph2 region that has been established though comparative genomic analysis with rice (Sutton et al., 2003).

Interestingly the mechanisms that restrict CO formation to homologs are not conserved between allopolyploids, each species having evolved its own mechanisms independently [reviewed in (Jenczewski and Alix, 2004)].

These mechanisms are not error free, evidence for CO between homoeologous chromosomes has been found in numerous species (Gaeta and Chris Pires, 2010; Chester et al., 2012a; Chalhoub et al., 2014; Lashermes et al., 2014; Li et al., 2015; Bertioli et al., 2016) although ongoing homoeologous exchanges are thought to occur only rarely (Sharpe et al., 1995). COs between homoeologs results in the formation of large Homoeologous Exchanges (HEs), the replacement of one chromosomal region (which is lost) with a duplicate of the corresponding homoeologous region (Nicolas et al., 2007; Gaeta and Chris Pires, 2010).

Suppression of CO between homoeologous chromosomes in allopolyploid does not mean however that the whole recombination process is abolished. The observation of synaptic multivalents at zygotene in several allopolyploids suggest that at least some early recombination intermediates have been formed between homeologous chromosomes. This is supported by the observation of early recombination nodules associated with the SC in wheat synaptic multivalents, which drop down dramatically as meiosis progresses (Hobolth, 1981). This suggests that early homeologous recombination intermediates are not committed to form CO but instead are resolved as non-crossovers. When this occurs, heteroduplexes may arise from sequence divergence at the site of strand invasion and, after resolution, result in non-crossover gene conversions between subgenomes. Accordingly, such very localized exchanges have been detected at the single-nucleotide scale in *B.napus, Coffea arabica* and allopolyploid cottons (Salmon et al., 2010; Flagel et al., 2012; Chalhoub et al., 2014; Lashermes et al., 2016). In *B.napus* (Chalhoub et al., 2014), reported that gene conversions explained 86% of the allelic differences between *B. napus* and its progenitor *B. rapa*. In cotton, the extent to which gene conversion between subgenomes contributes to genetic diversity is subject to controversy (Guo et al., 2014; Page et al., 2016).

1.6 Ploidy level and recombination frequencies

Interestingly, several studies pointed out that recombination frequency tends to increase as a consequence of a higher ploidy level. For example, comparative genetic mapping studies in allotetraploid cotton revealed that the At and Dt subgenomes experienced more than 50% higher recombination rates than their diploid counterparts (Desai et al., 2006). Likewise, in *Brassica*, almost all linkage groups of the A subgenome appeared to be longer in the allotetraploid *B.napus* than in the diploid *B. rapa* (Suwabe et al., 2008). More recently, Pecinka et al., 2011 confirmed that recombination frequencies increases in newly formed polyploids, whether they are autotetraploids (*A.thaliana* x *A.thaliana*) or allotetraploids (*A.thaliana* x *A. arenosa*), compared to diploid *A. thaliana*, all plants sharing an identical genetic background. This last result indicates that CO increase may occur irrespective of the nature (homologs / homoeologs) of the additional set of chromosomes.

The link between ploidy level and recombination frequencies has been more explicitly studied in Leflon et al., 2010 where the authors compared recombination frequencies between allotetraploid (AACC), triploid (AAC) or diploid (AA) *Brassica* hybrids sharing the same genetic background. They observed an increase in recombination frequencies in allotetraploid compared to diploid but also an unexpected boost in CO frequencies in allotriploid hybrids. I have addressed this issue with more details in paragraph 2.4 (see below p56.).

However, such cases should not be used to conclude that meiotic recombination is always highest in allotriploids. On the contrary, White and Jenkins (1988) and Jenkins and White (1988) observed that chiasma frequency was higher in *Scilla autumnalis* allotetraploid hybrids than in the corresponding allotriploid hybrid, which indicates that the observed increase is lineage specific.

1.7 Impact of recombination on genetic diversity

So far I have presented the mechanistic aspects of meiotic recombination, emphasising on CO formation and control. I will now review how the direct or indirect consequences of meiotic recombination can shape genome diversity and are thought to have major impact on plant genome evolution (Gaut et al., 2007).

In numerous plant species, genetic diversity correlates positively with local recombination rate (Roselius et al., 2005 and references within, Tenaillon et al., 2004; Wang et al., 2016). This observation has been interpreted as a direct or indirect consequence of recombination on polymorphism.

The direct effect of recombination can first result from the mutagenic nature of recombination itself. Irrespective of the final outcome (CO or NCO), all DSB repair mechanisms rely on the synthesis of short patch of DNA (Figure 6). Unlike DNA replication during the S phase of the cell cycle, DNA synthesis associated with DSB repair by homologous recombination is highly inaccurate (Malkova and Haber, 2012). In yeast, DSB repair during mitotic homologous recombination is accompanied by an increase in mutations near the site of the break. Recently, (Rattray et al., 2015) showed that this was also the case during meiotic recombination. These authors found a 6 to 21-fold increase in mutation rate after meiosis compared to the basal mutation rate observed after mitotic growth. This increase was dependent on SPO11, i.e., on the formation of DSBs, and was more pronounced when meiotic mutation rate was estimated close to a meiotic hotspot.

Another source of diversity directly linked to recombination is the generation of single-nucleotide mutations through GC-biased gene conversions (gBGC). gBGC can occur during the invasion step of meiotic recombination when parental alleles differ. It is hypothesized that, when the mismatch created in the heteroduplex is repaired, the changing of one of the nucleotides would slightly favour a conversion of an AT allele by a GC allele (Webster and Hurst, 2012). There is evidence for gBGC in yeast, mammals and birds and other species (Glemin 2016) but this is more equivocal in angiosperms. gBGC has been reported in rice (Muyle et al., 2011 but see Flowers et al., 2012) and in maize (Rodgers-Melnick et al., 2015) but not in. *A. thaliana*, where recombination positively correlates with AT-rich regions (Wijnker et al., 2013).

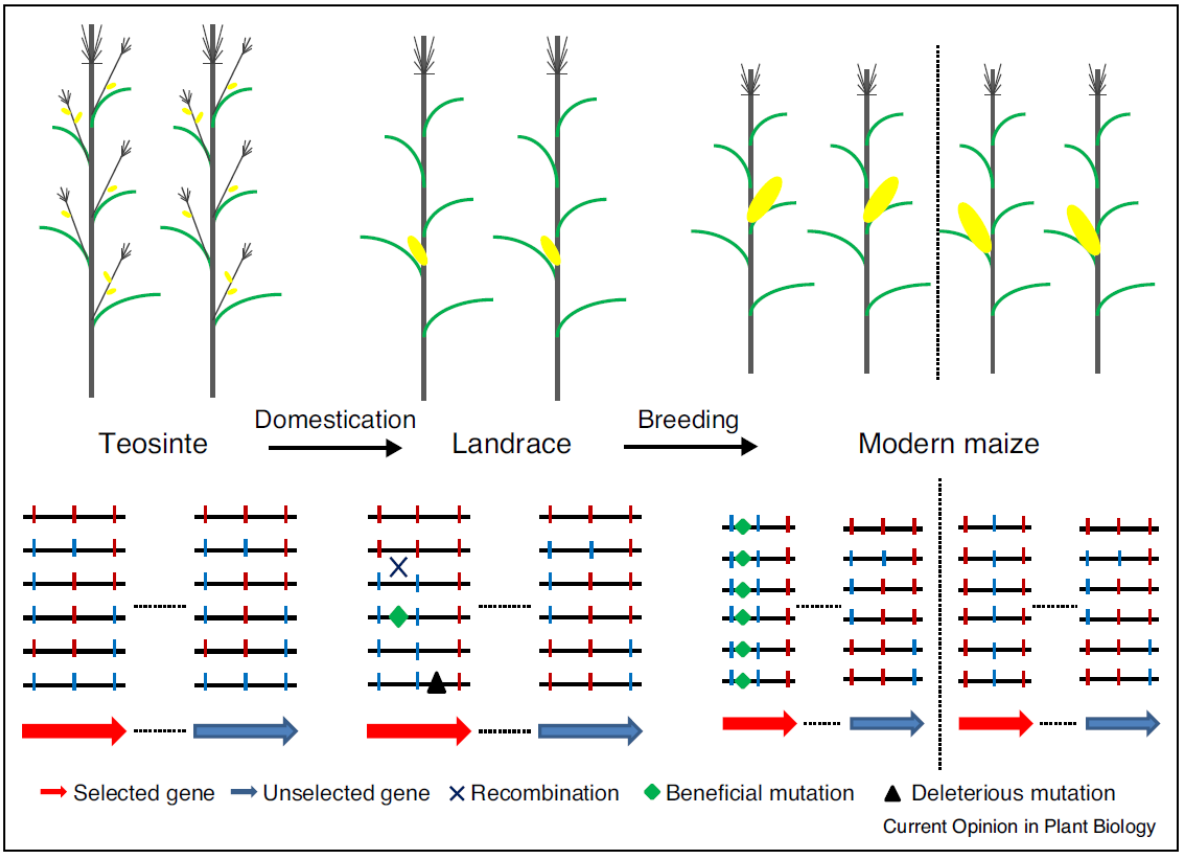


Figure 10 : Genomic changes with crop domestication and breeding [reproduced from (Shi and Lai, 2015)**]**
Upper plot: morphological changes during the domestication and breeding of maize. In Teosintes, the main stalk contained multiple branches ending in the tassel (inflorescence of male flowers) and bearing several small ears (female inflorescences) (yellow). After domestication, maize landraces retained only one primary branch with a moderate size ear along the stalk. Modern breeding generated maize cultivars with bigger ear.
Lower plot: genomic changes during domestication and subsequent breeding. In Teosintes, there are multiple haplotypes in both selected genes and unselected genes. After domestication, the number of haplotypes are reduced due to genetic bottleneck, the decrease is more pronounced in selected genes.
In landrace, a beneficial mutation (green diamond) and a recombination which generated a new haplotype (red–blue–red) are selected. Concurently, a deleterious mutation (black triangle) occurred in a selected gene. As a result of positive sweeps and background selection, only the newly generated haplotype (red–blue–red) and the haplotype with the beneficial mutation (blue–green–blue–red) are retained. Diversity is preserved in non-selected genes.

If the process of recombination itself was mutagenic and neutral, one would expect to find a correlation between recombination intensity and species divergence. However such a link has not been established when assessed in tomato (Roselius et al., 2005), poplar (Wang et al., 2016b) or *Z. mays* in (Tenaillon, 2001; Tenaillon et al., 2002). Glémin, 2010 suggests however that gBGC may have non-negligible fitness consequences when taking into account the indirect effects of meiotic recombination.

Irrespective of whether the direct consequences of recombination can account for a significant amount of polymorphism in different genomic regions, meiotic recombination influences the evolutionary fate of a mutation through its effect on natural selection (Figure 10). Positive selection for an allele increases the frequency of this allele in the population but also the frequencies of genetically linked alleles which are dragged along in a "positive sweep". The lower the recombination frequency, the larger the region that is swept; i.e., where genetic diversity is erased as a consequence of local selection and linkage disequilibrium (association of alleles at different loci). This "hitchhiking effect" can also occur when detrimental alleles are selected against (''background selection''). The reduction in nucleotide variability associated with selection may not be restricted to regions of suppressed recombination; it may also be apparent in any region where the density of selected mutations is high relative to local rate of recombination. In *A. thaliana* and rice for example (Nordborg et al., 2005; Flowers et al., 2012), gene density is a better predictor of the level of polymorphism than recombination rate. In addition, the impact of selection on genetic variation varies depending on the species. As selfing reduces the effective recombination rate, "hitchhiking effect" could possibly be more pronounced in partially self-fertilizing plants than in outcrossers (Nordborg, 2000).

As a result, low recombining regions usually display a low polymorphism and tend to accumulate deleterious mutations as a consequence of inefficient selection. This is for example illustrated in maize where deleterious polymorphism is less frequently found within areas of high recombination (Rodgers-Melnick et al., 2015).

1.8 CO frequencies and selection

It remains unclear why natural variation for CO frequencies is kept so low in the vast majority of species. Recombination rate is subject to selection. This seems also to be the case for the direction and level of interference, as seen in Drosophila (Aggarwal et al., 2015) and in maize (Bauer et al., 2013). Modifying recombination rate is indeed not neutral. On one hand, COs reshuffle the genome; they contribute to create new combinations of alleles that may result in novel phenotypes or in new epistatic interactions; in turns, these genetic novelties may affect the organism's fitness and its ability to respond to selection. On the other hand, COs may also break existing positive associations of alleles, thereby setting a threshold above which higher recombination would be selected against. The optimal recombination rate would then be variable; higher recombination rate would be favourable only in certain conditions where new genetic diversity would bring fitness advantages.

Both theory and simulations show that selection generally favours an increased recombination rate during periods of rapid evolutionary change (Otto and Barton, 1997). Natural variation for recombination frequencies would be at use in the context of strong directional selection imposed over multiple loci when genetic variability is limited by linkage disequilibrium. Numerous studies have shown that recombination frequency and sometimes CO interference were modified after strong artificial selection for other characteristics (Otto and Lenormand, 2002 and references within, Aggarwal et al., 2015). In yeast, it has been observed that sex increases adaptation rate to a new harsh environment but had no measurable effect on fitness in a new benign environment where there is little selection (Goddard et al., 2005).

This relationship between selection and increased recombination frequencies is however less clear when considering domestication, a somewhat slower adaptation process (Purugganan and Fuller, 2009). For example, in mammals, it was shown that domestication did not result in an increased recombination rate contrary to plants (Munoz-Fuentes et al., 2015).

How limiting are (low) recombination frequencies for breeding effort is actually an open question in the literature. Only a few studies have investigated what would be the gain in term of breeding efficiency if higher recombination rates would be achieved.

Studies in soybean (Piper and Fehr, 1987) and wheat (Altman and Busch, 1984) assessed the advantages of including generations of intermating to increase recombination in the parental population before applying selection for agronomical traits. They observed limited to no changes in the means of the resulting selected population. (Melchinger et al., 2003) came to the same conclusions in corn although they observed an increase in transgressive phenotypes in their hybrid populations. This seems positive at first as a larger utilisable genetic variance was released. However, the authors estimated that the odd of recovering better lines in the intermated populations was rather low. For their trait of interest (grain yield), they concluded that the disruption of beneficial gene combinations that already exist in elite cultivars greatly outweighed the advantages of increasing recombination.

A major limitation to those studies however is that they were unable to measure the extent to which their breeding scheme resulted in increased recombination frequencies. Moreover more recent studies in soybean revealed that genotypes resulting from crop improvement showed no decrease in recombination frequencies (Pfeiffer, 1993) and even tended to result from more recombination events than unselected lines in the same population (Stefaniak et al., 2006). Only a few simulation studies have specifically tested whether higher CO frequencies would positively affect selection efficacy. (McClosky and Tanksley, 2013) found only relatively modest gains (11%) in response to selection. Another simulation in livestock suggests that substantial increase in gains in response to selection would require a large increase in recombination frequencies (33% gain obtained with a 20-fold increase) (Battagin et al., 2016). In all these studies however, for sake of simplicity, the simulated recombination rates were computed without taking in account interference. This thus inflated the number of COs predicted to occur in the wild type and reduced the output gain of increased recombination. Finally, in Kessner and Novembre, 2015, the authors emphasise the importance of recombination when conducting artificial selection experiments to detect and localize QTL contributing to a quantitative trait.

1.9 How to tackle some of the breeder's challenges?

Some of the aspects of meiotic recombination discussed above, especially the various regulatory mechanisms responsible for CO patterning, can represent obstacles for crop improvement. An increasing number of studies have, over the years, addressed how fundamental knowledge gained about the underlying mechanisms of meiotic CO formation could be of interest for plant breeding purposes.

- As mentioned above, recombination frequencies are on average relatively low in most species. This could represent a limiting factor when looking for new combinations of beneficial alleles in a progeny or introducing a trait of interest into an elite genotype. Low CO frequencies also reduce the odds to remove linkage drags between deleterious and beneficial genes and limit the power of mapping and positional cloning approaches.

  The recent characterization in *A. thaliana* of multiple pathways that limit class II CO frequencies (FIGL, FANCM, RECQ see Figure 4) may offer a way to release the constraint of having only a few COs per chromosome. Indeed the effect of both *figl1*, *top3α-R640X* and *recq4a recq4b* mutations on CO formation was shown to be cumulative with *fancm* leading to a sixfold and a ninefold increase in CO frequency respectively without immediate negative effects on meiosis and fertility (Séguéla-Arnaud et al., 2015). An alternative to using mutants is to look for natural variation for CO patterning. Although evidence for cis and trans natural variation for CO frequencies have been found in several species, notably in *A. thaliana* (López et al., 2012; Sanchez-Moran et al., 2002, Esch et al., 2007), *Zea mays* (Dole and Weber, 2007; Esch et al., 2007; Timmermans et al., 1997; Yandeau-Nelson et al., 2005), and wheat (Esch et al., 2007), so far no recombination modifiers have been characterized in plants. In animals, genome wide associations studies have repeatedly associated heritable variation in recombination frequencies with an handful of loci, notably RNF212, CPLX1, REC8 and PRDM9 (Kong et al., 2008; Chowdhury et al., 2009; Sandor et al., 2012; Reynolds et al., 2013; Kong et al., 2014; Ma et al., 2015) in cattle, human and mouse. In all case, all of these recombination modifiers had a relatively small effect (between 1 and 2 fold).

- Recombination frequencies are not homogeneous along the chromosomes; I mentioned above that cold regions alternate with hot regions for CO frequencies. Although cold regions are often heterochromatic, this does not mean that they are deprived from genes that can be of interest for the breeders as observed in tomato, barley, maize and wheat for example (Sato et al., 2012; International Barley Genome Sequencing et al., 2012; McMullen et al., 2009; Choulet et al., 2014). In barley and maize, low recombination centromeric and pericentromeric regions contain around 20 % of total gene content. This is more extreme on chromosome 3B of wheat where 70 % of total gene content is found in CO poor regions (Choulet et al., 2014). As a result, the confidence intervals of QTLs in regions where recombination is suppressed regions can cover hundreds of megabases. Thus, fine mapping strategy cannot be undertaken for these QTLs.

  Because of low recombination frequencies, polymorphism is expected to be low in CO-poor regions. Thus, although challenging, there may be an interest to specifically increase CO frequencies in these regions; this would not only allow to increase the amount of diversity in these regions but also to efficiently purge the deleterious alleles that tend to accumulate there. Although promising, the burst in recombination frequencies described above might not hold true for all genomic regions. As observed in Girard et al. (2015), the increase in CO formation in *figl1* mutant is more pronounced in distal regions of the chromosomes. In other words, CO frequencies but not CO localization is affected in *figl1*.

  Unlike for CO frequencies, only one genetic determinant for CO localization has been identified so far. In Jahns et al. (2014), CO localization but not CO frequencies is modified in the *axr1* mutant in *A. thaliana* leading to clusters of class I CO. This CO clustering was however not observed in centromeric regions.

Another strategy has modified chromosome primary structure in order to play around with the centromere-telomere gradient for CO frequencies (see above). Qi et al. (2002) managed to increase CO frequencies in proximal, usually low-recombining regions by placing them close to the chromosome end. In a different strategy, Ederveen et al. (2015) used pollen irradiation in *A.thaliana* to generate large structural variation (deletion and inversion) where meiotic recombination cannot occur. In most case CO homoeostasis resulted in an increase of CO frequencies in regions proximal to the structural variant. Although the largest increase was observed in regions close to the telomere, they nonetheless noted a maximum increase just over 150% of CO frequencies in intervals proximal to the centromere.

Under the hypothesis that the lack of DSBs is limiting for CO formation in these regions, a promising approach is the targeting of DSBs to specific sites on the genome as it has been achieved in yeast (Peciña et al., 2002). Although there is no assurance that a DSB will necessary yield a CO, several approaches could be tempted in plants to specifically induced DSB formation using SPO11 fusions with a variety of different DNA-binding domains (Nogué et al., 2016).

- Breeders often use related species to introduce genetic diversity in their population. However, meiotic recombination is sensitive to sequence divergence and mechanisms exist that restrict recombination to homologous over non-homologous chromosomes (see above).

  The use of mutants for loci that limit homoeologous recombination has been successfully used for introgression purposes in wheat (Rey et al. (2015) and references within). For example, Lukaszewski, (2000) used *ph1* mutants to eliminate the quality defect associated with an introgression coming from rye (*Secale cereale* L.). The use of okadaic acid to phenocopy the *ph1* mutant effect could provide a way to reversibly allow non homologous recombination and thus to maintain fertility and genome stability once the desired recombinants are obtained (Knight et al., 2010).

- In many species, the breeding process is performed on parental lines that are fixed for a number of traits of interest. These parental lines are then crossed in a controlled manner to produce heterozygous F1 hybrids that combine the desired traits and also benefits from hybrid vigour. The production of F1 seeds thus requires a constant fresh production of elite hybrids and the resulting heterozygotes cannot be used as a basis for a new breeding program. Different approaches have been proposed to achieve preservation of the elite heterozygotes genotype either as a mean to simplify the hybrid production process (Bicknell and Koltunow, 2004) and/or as a basis for further crop improvement.

  One of these approaches is apomixis (asexual formation of a seed from maternal material) that occurs in a wide range of species but rarely in crops. A recent proof of principle has illustrated the potential use of apomixis in plant breeding and seed production by demonstrating that complex characters could be stably inherited across generations in a natural apomictic hybrids (Sailer et al., 2016). Apomixis is under the control of a limited number of loci but the corresponding genes have not yet been identified (Pupilli and Barcaccia, 2012; Koltunow et al., 2011). So far, the attempts to directly introgress apomixis into crops have been unsuccessful. However, it is possible to engineer apomixis de novo. Briefly, this has been achieved by turning meiosis into mitosis through a disruption of meiotic recombination, homologous chromosome segregation and cell cycle. Concretely this required to combine mutations for SPO11 (recombination), REC8 (segregation) and OSD1 (to avoid a second round of cell division) into a single genotype called MiMe in *A. thaliana* (D'Erfurth et al., 2009). Further seed production from a fixed MiMe hybrid is however problematic because selfing would lead to doubling of ploidy at each generation. A way to tackle this issue has been to cross MiMe plants with a GEM (genome elimination) line whose genome is eliminated post-fertilization when crossed with any other genotype (Ravi and Chan, 2010). This step is nonetheless limiting for clonal seed production because of its relatively low efficiency (Marimuthu et al., 2011).

A second approach is reverse breeding, a genetic engineering process based on the suppression of crossover formation in a hybrid plant of interest (Dirks et al., 2009). Unrecombined chromosomes (identical to the parents) are left free to segregate randomly to daughter cells during the first meiotic division. The few viable spores that combine by chance one copy of each chromosome are then regenerated via double haploids production. The outcome of reverse breeding is then a set of substitution lines that contain a varying number of unrecombined paternal or maternal chromosomes due to random chromosome segregation. Plants identical to the parents can then be used to reproduce the hybrid and/or as a basis for new breeding programs while the set of substitution lines is extremely valuable for mapping quantitative trait loci (QTL) and for advanced forms of marker assisted breeding. While the proof of concept has been published in *A. thaliana* (Wijnker et al., 2012), one major limitation of this technique is the random segregation of chromosomes during the first meiotic division which makes it difficult to adopt in crops with high chromosome numbers as well as in polyploids.

Chapter 2: The plant model

In this chapter I will briefly introduce *Brassica napus* and give the necessary background information for the comprehension of my work. I will notably review the polyploid origins of *Brassica napus* and its position within the *Brassicaceae*, the relevance of this family for plant breeding and what makes *Brassica napus* a model to study natural variation of CO frequencies between homologous and homoeologous chromosomes.

2.1 Origin of *Brassica napus*

Rapeseed (*Brassica napus*, AACC; 2n=38) is a member of the large *Brassicaceae* family (~325 genera and 3,740 species [reviewed in (Hohmann et al., 2015)] that include various crops and the model species *Arabidopsis thaliana*. Rapeseed is a recent allopolyploid species that formed from hybridization events between the ancestors of modern *B. oleracea* (CC; 2n=18) and *B. rapa* (AA; 2n=20); these two diploid species diverged from a common ancestor less than 4 million years ago and their genomes were brought back together only recently to form *B. napus* (around 7500 - 12500 years ago (Chalhoub et al., 2014). As no truly wild *B. napus* population has been reported, hybridisation between *B. napus* progenitors is thought to have occurred in cultivated contexts, as a result of either accidental or deliberate inter-specific crosses between crops that were cultivated alongside. The original hybridisation events that gave rise to *B. napus* occurred more than once, and involved different maternal genotypes that are probably related to *B. rapa* or an A genome relative (Allender and King, 2010). Genetic diversity analyses revealed a strong population structure, mainly explained by growth habits (spring or winter) and geographical origin (Asian or European for winter types) (Gazave et al., 2016).

The recent release of reference genomes for *Brassica napus* (Chalhoub et al., 2014), *B. rapa* (Wang et al., 2011) and *B. oleracea* (Liu et al., 2014; Parkin et al., 2014) provided further insights into the dynamics of *Brassica* genome evolution and divergence. The *B. napus* genome is around 1,2Gb in length (Arumuganathan and Earle, 1991) and contains a minimum of 101,040 gene models (Chalhoub et al., 2014). The assembled C_n subgenome (525.8 Mb) is larger than the A_n subgenome (314.2 Mb) (Chalhoub et al., 2014). This is consistent with the relative sizes of the assembled C_o genome of *B. oleracea* (~630 Mb) (Liu et al., 2014; Parkin et al., 2014) compared to the A_r genome of *B. rapa* (312 Mb) (Wang et al., 2011).

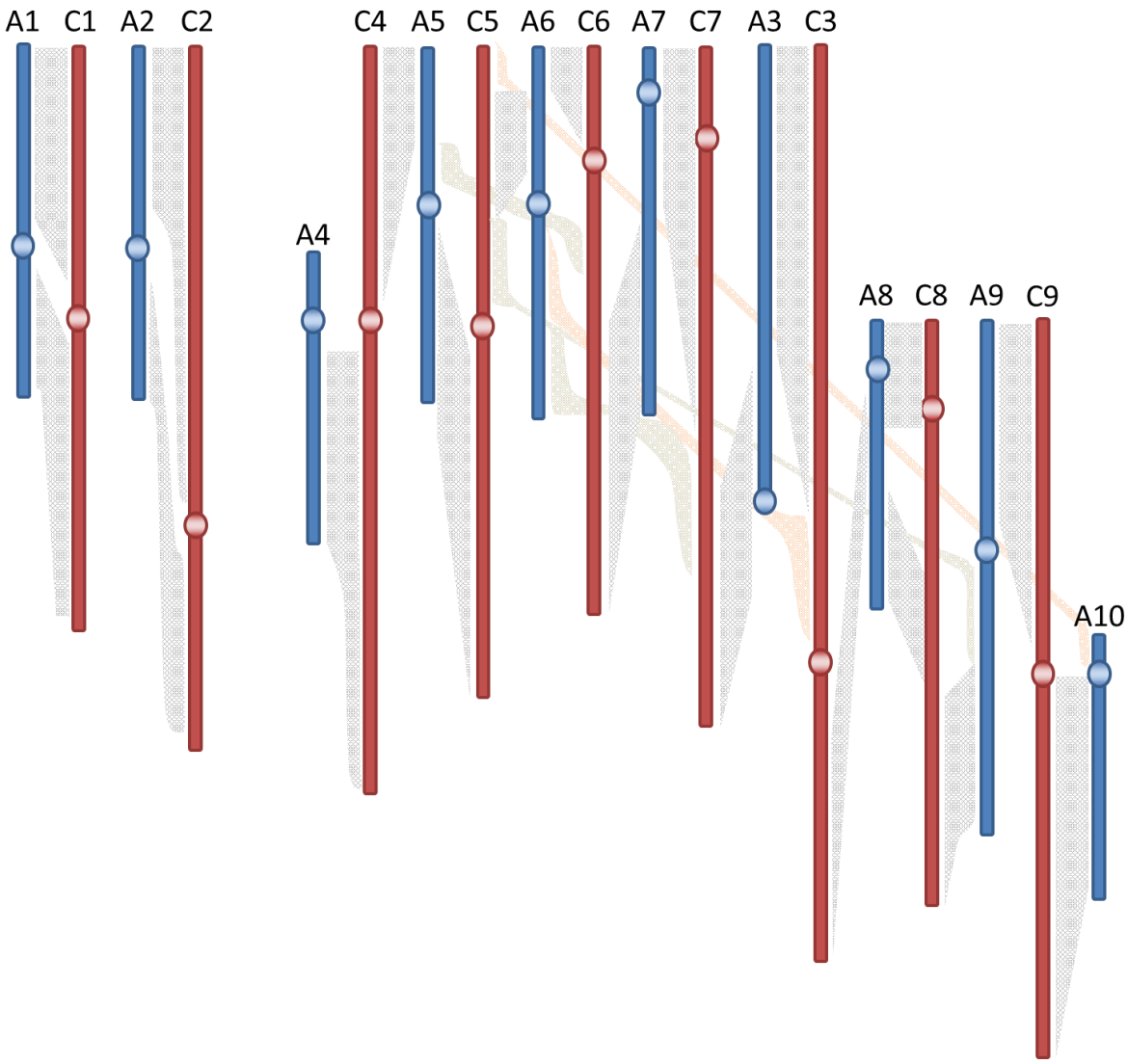


Figure 11: Collinearity of A and C subgenomes

The 10 A (blue) and 9 C (red) chromosomes of the genome of *B. napus* are arranged so that extensive collinearity (represented by hatching) between chromosome arms becomes apparent.

The difference in size between *B. oleracea* and *B. rapa* results from multiplication of both repetitive and genic sequences (45,758 in *B. oleracea* vs 41,174 in *B. rapa*). This is reflected in *B. napus* genome, which displays an asymmetric distribution of genes and transposable elements between subgenomes (Chalhoub et al., 2014). This notwithstanding, most orthologous gene pairs in *B. rapa* and *B. oleracea* have remained as homoeologous pairs in *B. napus* (Chalhoub et al., 2014), with only limited gene loss after *B. napus* formation.

The A and C genomes are composed of 2n=20 and 2n=18 chromosomes, respectively. Although A and C chromosomes are extensively collinear (Figure 11), this collinearity does not usually extend more than one chromosome arm, the second arm of each chromosome being collinear to another homoeologue (Figure 11). Only the A1-C1 and A2-C2 homoeologous pairs are collinear along their entire length. In a few cases (A6 and C5 notably), more complex rearrangements differentiate the A and C chromosomes. Comparison of sequence identity between collinear regions revealed a 8.4 and a 5.7% InDel and SNPs difference, respectively (Ming and Man Wai, 2015). Cheung et al., 2009 estimated that transcripts in homoeologous pairs differ in sequence, on average, by approximately 3.5% while Higgins et al., 2012a found a mean density of SNPs between homoeologs gene pairs of ~1 %.

Brassica genomes have undergone an extra whole genome triplication (WGT) event compared to *A. thaliana* (Figure 12). This WGT is thought to have occurred in two steps; first a tetraploidization event with two diploid genomes (MF1 and MF2) and then a second polyploidization event involving a third diploid genome (LF) (Figure 13). As a consequence, each genomic region in *A. thaliana* corresponds to six genomic regions in *B. napus* and each gene in *Arabidopsis thaliana* has up to 6 homologs in *Brassica napus,* termed paleologs. Such a high number of homologs is rarely observed because the additional copies of a gene resulting from polyploidization tend to be lost over time. This process termed fractionation (Freeling, 2009; Woodhouse et al., 2010) can be biased, one subgenome retaining more genes compared with the other. For example, in *B.rapa* the LF subgenome retained 70% of the genes found in *A. thaliana* while this proportion is less for the MF1 and MF2 sub-genomes (46% and 36%, respectively (Wang et al., 2011). (Lloyd et al., 2014) showed that fractionation follows a predictable pattern in a wide range of species (14 polyploidization events ranging in age from 5–9 to approximately 130My).

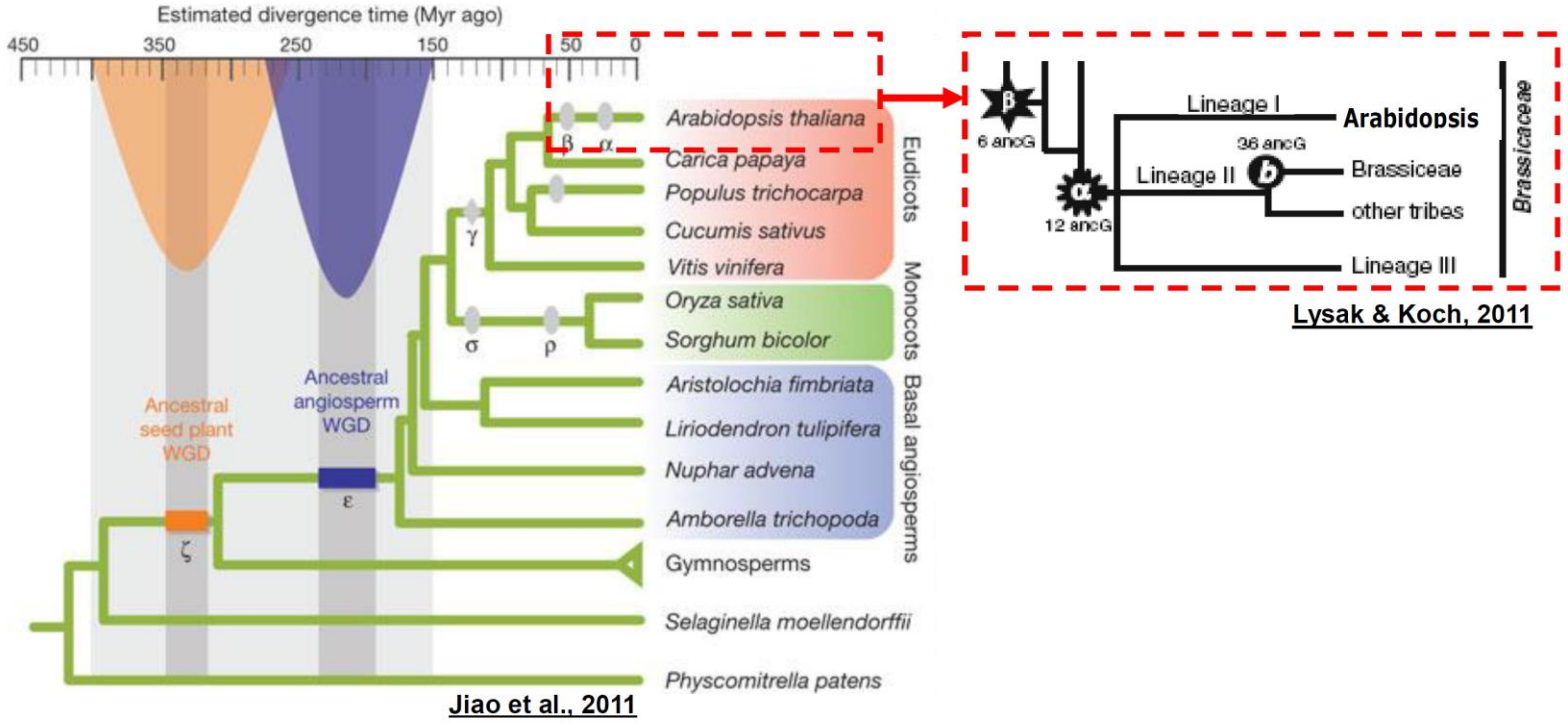


Figure 12: The course of polyploidy events from ancestral angiosperm to Brassica species [Figure and text reproduced from (Liu et al., 2014)]

From the ancestral species of all extant seed plants to *Brassica* lineage, there were at least six polyploidy events undergone: ζ occurred in an ancestral species of all extant seed plants, ε in angiosperm plants, γ in core eudicots, β in *Brassicales*, α in *Brassicaceae*, and another *Brassica* lineage-specific whole genome triplication, b, after split from Arabidopsis.

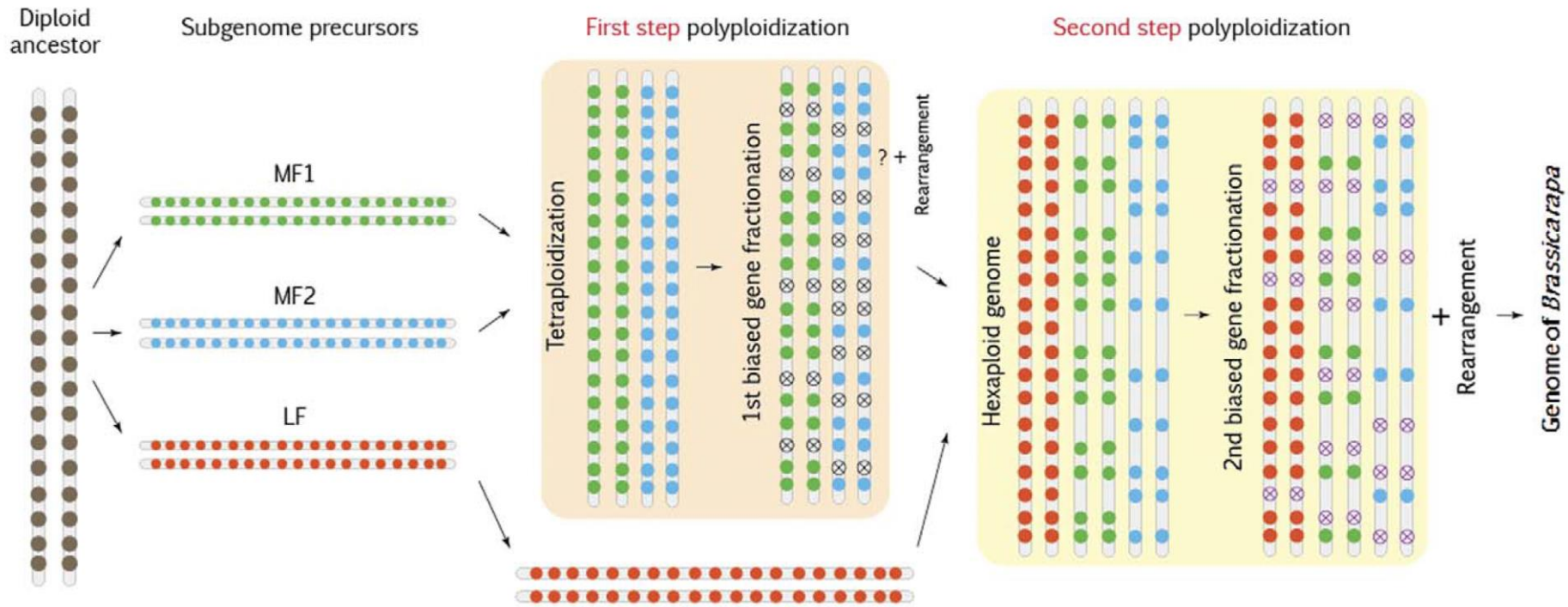


Figure 13 : Flow chart of the ‘‘two-step theory’’ to explain the genome triplication that occurred in the early stages of the origin of *B. rapa* species [Figure and text reproduced from (Cheng et al., 2012)]

Circles denote genes and circles with crosses indicate genes that are not detectable. Red circles are genes in subgenome LF, blue and green circles are genes in subgenomes MF1 and MF2, respectively.

Duplicate loss is maximal right after the onset of polyploidization and progressively slows through time until eventually reaching a plateau for very old WGD [see also (Sankoff et al., 2010; Li et al., 2016)]. Meiotic genes and more specifically meiotic recombination genes display a similar trend, although more pronounced. There is no indication however that the loss of meiotic genes duplicates is subject to selection; Lloyd et al., (2014) thus proposed that this fate actually reflects what happens when there is no (or little) selective force opposing duplicate loss and that the higher retention of duplicates observed genome wide would result from the inclusion of genes selectively maintained in duplicate (e.g. dosage-sensitive genes; (Lloyd et al., 2014).

Along with fractionation, a mechanistically distinct form of gene loss results from segregation of large homoeologous exchanges (HEs) (Chalhoub et al., 2014). A will be seen; I have been incidentally introduced to these events during my PhD work. All details are provided in Chapter 4 (p.61).

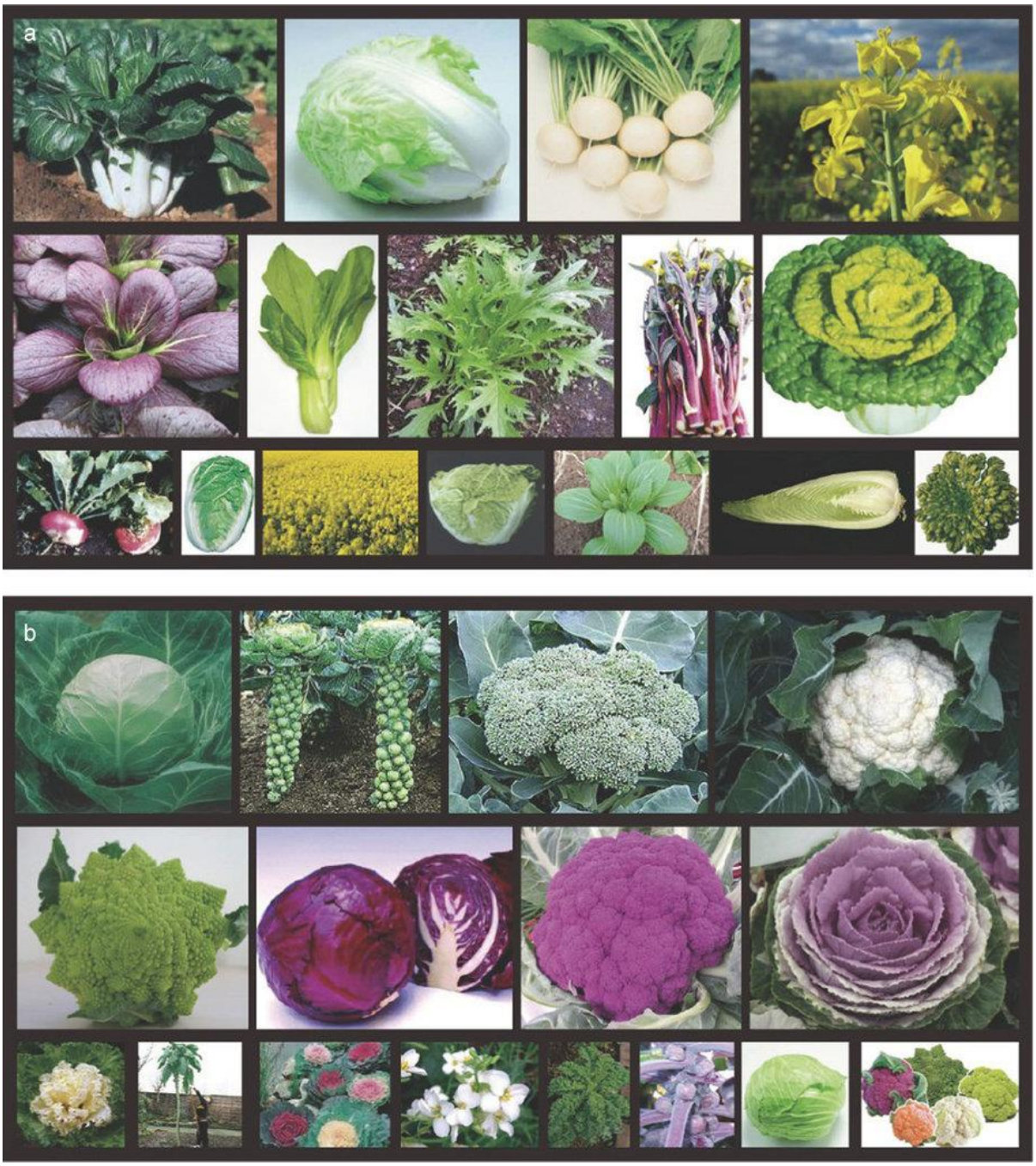


Figure 14 : Rich morphotypes of *Brassica* plants [Figure and text reproduced from (Cheng et al., 2014)**]**

(**a**) Morphotypes of *B. rapa*; top two lines from left to right: pak choi, heading *B. rapa*, turnip, oilseed, purple pak choi, caixin, mizuna, purple caitai and takucai; the third line shows additional morphotypes or varieties of the previous morphotypes. (**b**) Morphotypes of *B. oleracea*; top two lines from left to right: heading cabbage, Brussels sprouts, broccoli, cauliflower, purple cabbage, purple cauliflower, collard; the third line shows additional morphotypes or varieties. Some of the pictures were collected from the Internet.

2.2 Relevance for breeding

Each Brassica crop (*B. rapa*, *B. oleracea* and *B. napus*) shows a rich diversity of morphotypes including leafy heads (Chinese cabbage [AA], cabbage [CC]), enlarged roots (turnip [AA], rutabaga [AACC]), other enlarged organs like stems and inflorescences (cauliflower, Brussel sprouts [CC]), oilseeds (both AA and AACC (Figure 14). Although any of these species can be used as either a vegetable, fodder, oilseed or even as ornamental crop, *Brassica rapa* and *Brassica oleracea* are often referred to as leaf vegetables and *Brassica napus* as an oilseed crop.

Most of the breeding efforts for rapeseed have been dedicated to increase seed yield and to reduce the content of nutritionally undesirable components of the oil and of the seed hull. These efforts led to the development of the double low ("00") varieties that display concurrently low erucic acid content, which is undesirable in edible oils, and low Glucosinolates (GSLs) content, which in animal feed can result in goitrogen-induced hypertrophy. Among the other objectives currently followed by breeders, a lot of effort has been also invested in the development of "yellow seeded" varieties resulting from reduced condensed tannins content and associated with higher oil and protein content and lower fibber content. Development of varieties with oil properties meeting the requirement of the food processing industry (high oleic and low linolenic acid content) or more recently the development of oils suitable for conversion to biodiesel and industrial lubricants is also a recurrent plant breeding objective, along with the identification of genotypes able to grow under low input farming regimes (especially low nitrogen input).

The narrow origin of B*rassica napus* associated with intense selection has resulted in a notable decline in genetic diversity in modern cultivars. Most of the current crop germplasms are related (Hasan et al., 2006; Qian et al., 2014) and a strong deficit in polymorphism is observed in regions where QTLs for GLS and erucid acid were mapped (Qian et al., 2014). Although breeders attempted to reintroduce diversity through introgression from *B. rapa* and *B. oleracea*, as well as other related *Brassica* species, they focussed their efforts on a few phenotypic traits of interest. Loss in genetic diversity is more pronounced for the C genome (Qian et al., 2014) probably because of less interspecific hybridization.

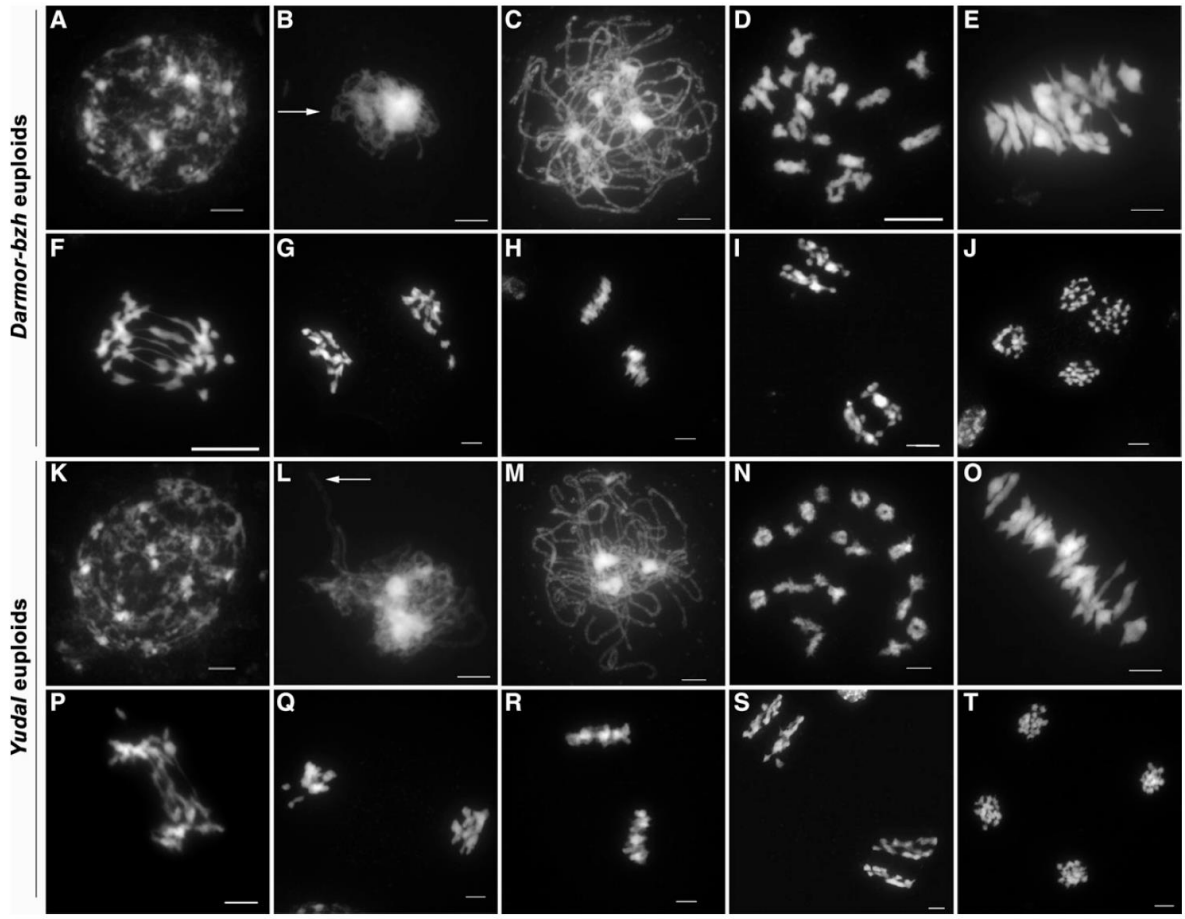


Figure 15. Meiosis in Euploid *B. napus* [Figure and text reproduced from (Grandont et al., 2014)]

DAPI staining of pollen mother cells during meiosis of *Darmor-bzh* ([A] to [J]) and *Yudal* ([K] to [T]) euploids. Leptotene ([A] and [K]): Chromosomes condense and become visible as unpaired threads. Zygotene ([B] and [L]): Arrows indicate several close juxtapositions of chromosomes that probably mark the initiation of the synaptonemal complex. Pachytene ([C] and [M]): All chromosomes are closely aligned with one another, suggesting that the synaptonemal complex is complete. Diakinesis ([D] and [N]): chromosomes are condensed and form discrete separate bivalents. Metaphase I ([E] and [O]): All bivalents aligned on the metaphase plate. Anaphase I ([F] and [P]): Homologous chromosomes, each composed of two sister chromatids, move to the opposite poles. Telophase I ([G] and [Q]). Metaphase II ([H] and [R]). Anaphase II ([I] and [S]): Individual chromatids segregate to the spindle poles. Late anaphase II ([J] and [T]): Cells contain four sets of 19 chromatids. Bars = 5 mm

2.3 Meiosis in *Brassica napus*

Natural and resynthesized *B. napus* display very different levels of meiosis regularity and genome stability (for review, see Gaeta and Chris Pires, (2010); Jenczewski, (2013)). Although chiasmatic associations between A and C chromosomes are commonplace in synthetic *B. napus* (30 – 47.5% of cells; Szadkowski et al., 2010), meiosis in "natural" *Brassica napus* displays a diploid like behaviour, the formation of CO being to a large extent restricted to homologous chromosomes (Figure 15). This difference suggests that, like wheat, natural *B. napus* has evolved or inherited Pairing homoeologous loci that ensure proper chromosome recombination and segregation.

A detailed cytological characterization of meiosis in two genotypes (*Darmor-bzh*, a French dwarf winter cultivar and *Yudal*, a spring korean line) representative of the two main *B.napus* gene pools (Harper et al., 2012) has recently confirmed that *B. napus* displays a diploid-like meiotic behaviour (see Figure 2 and Figure 15) . Homologous and homoeologous chromosomes are sorted early on during prophase I; no more than one or two synaptic quadrivalents (association of four chromosomes joined at different points) are observed in ~50% of the cells at late zygotene – pachytene and are eliminated before diakinesis (Grandont et al., 2014). The mechanism(s) responsible for this early chromosome sorting is (are) unknown but appear(s) to be equally efficient in the two genotypes. This notwithstanding, an earlier channelling of recombination intermediates into the CO pathways was observed in *Yudal* compared to *Darmor-bzh* (Grandont et al., 2014). The two varieties not only differ in the progression of recombination but also in the number of class I CO, which is reduced in *Yudal* compared to *Darmor-bzh.* These differences between genotypes were not due to obvious differences in sister chromatid cohesion or the assembly of meiosis-specific chromosome axes, which are correctly formed both in *Darmor-bzh* and *Yudal.* Full synapsis occurs in both genotypes although it appears more diffuse in *Yudal* compared to *Darmor-bzh* (Grandont et al., 2014).

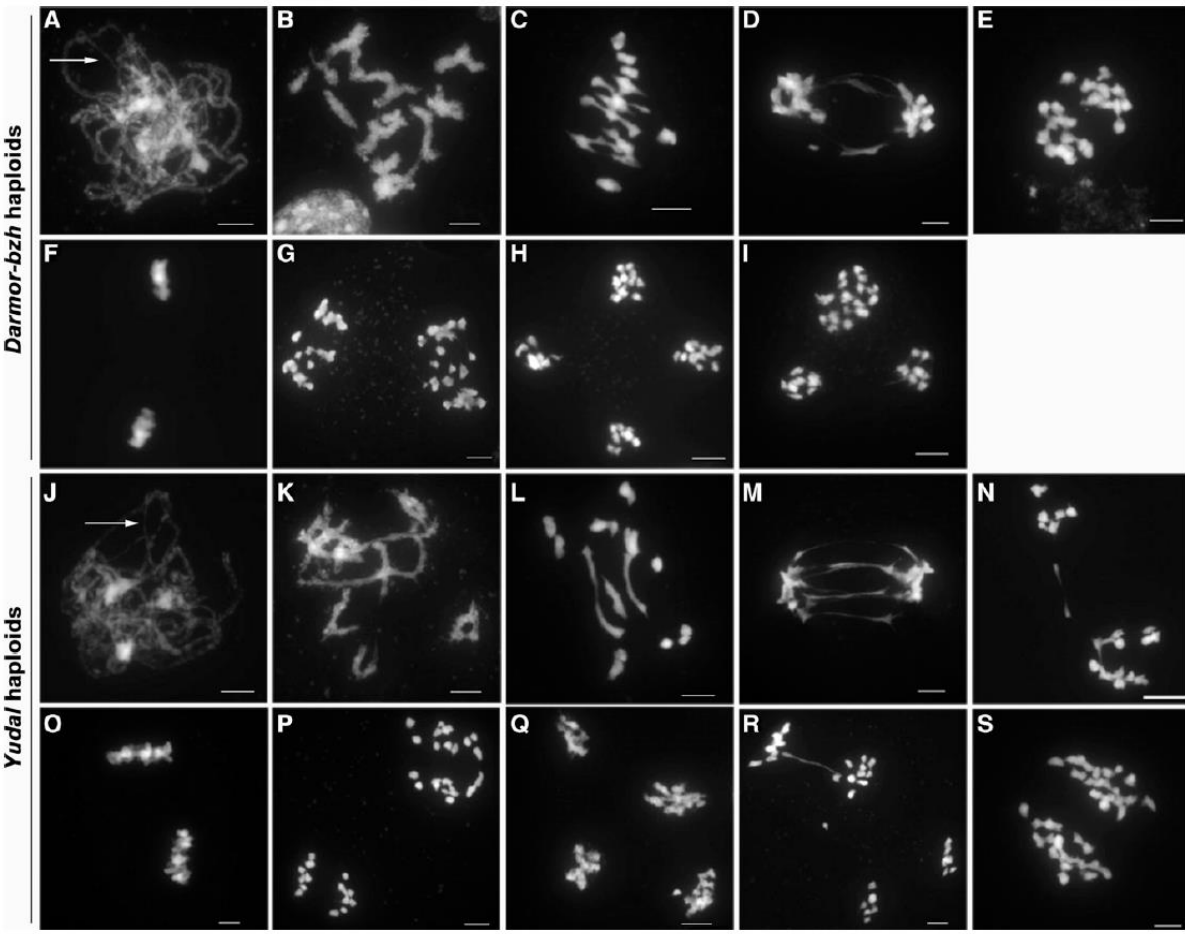


Figure 16. Meiosis in Allohaploid *B. napus* [Figure and text reproduced from (Grandont et al., 2014)]

DAPI staining of pollen mother cells during meiosis of *Darmor-bzh* ([A] to [I]) and *Yudal* ([J] to [S]) allohaploids. Pachytene ([A] and [J]): The presence of chromosomes that are not closely juxtaposed with one another (arrows) suggests that the synaptonemal complex is incomplete. Diakinesis ([B] and [K]). Metaphase I ([C] and [L]) with variable numbers of bivalents and univalents. Anaphase I ([D] and [M]): Non homologous chromosomes, each composed of two sister chromatids, are separated. Telophase I ([E] and [N]): Two groups of chromosomes are observed, indicating that univalents moved to one or the other pole of the cell. Metaphase II ([F] and [O]). Anaphase II ([G] and [P]): Individual chromatids segregated to the spindle poles resulting in the formation of different kinds of meiotic products, including unbalanced tetrads ([H], [Q] and [R]), triads (I), as well as dyads (S). Bars = 5 mm.

Although efficient, the process of homeologous chromosome sorting is not completely error proof as rare homoeologous exchanges can be evidenced in modern *B.napus* (Sharpe et al., 1995; see Chapter 4, p61). Interestingly, these events become dominant in allohaploids (AC), the meiosis of which was observed in detail by Grandont et al. (2014). The early stages of meiosis in allohaploid plants are similar to those described in euploids with no apparent defect in sister chromatin cohesion and chromosome axes. The first noticeable difference between euploids and allohaploids meisois occurs at pachytene when synapsis between pairs of non-homologous chromosomes is never completed (Figure 16). At metaphase I, both univalents (i.e., chromosomes that failed to form chiasmata) and bivalents are observed; I these plants, chiasmata are necessarily formed between non-homologous A and C chromosomes because allohaploids do not contain homologous chromosomes. In subsequent stages, chromosome segregation is irregular and leads to unbalanced tetrads and unviable pollen grains. Only a few microspores can eventually inherit the 19 chromosomes of the basic *B. napus* chromosome set (Grandont et al., 2014); they generate viable "unreduced" gametes that can be used to produce progenies (Nicolas et al., 2007, 2009, 2012).

The number of CO that are formed between non homologous chromosomes strongly differs between *Darmor-bzh* and *Yudal* with twice as many chiasmata observed in *Darmor-bzh* than in *Yudal* (Grandont et al., 2014). Genetic (Nicolas et al., 2009) and cytological (Grandont et al., 2014) assays indicate that all chromosomes are intrinsically able to form CO in each of the two varieties (at the haploid stage); however, the odds of forming a bivalent for a given pair of homoeologs varies depending on the genotype, sometimes in the opposite direction to what is expected (e.g. the A3-C3 is observed more often in *Yudal* than in *Darmor-bzh* allohaploids). This being said, the two clear cut phenotypes described in *Darmor-bzh* and in *Yudal* seems representative of the whole range of meiotic phenotypes that can be observed within B. napus (Cifuentes et al., 2010). High ("*Darmor-bzh* like") or low ("*Yudal* like") CO frequencies between homoeologous chromosomes have been repeatedly observed in allohaploids produced from a subset of varieties representative of *B.napus* genetic diversity, with only slight variations within these phenotypes (Figure 17) (Cifuentes et al., 2010).

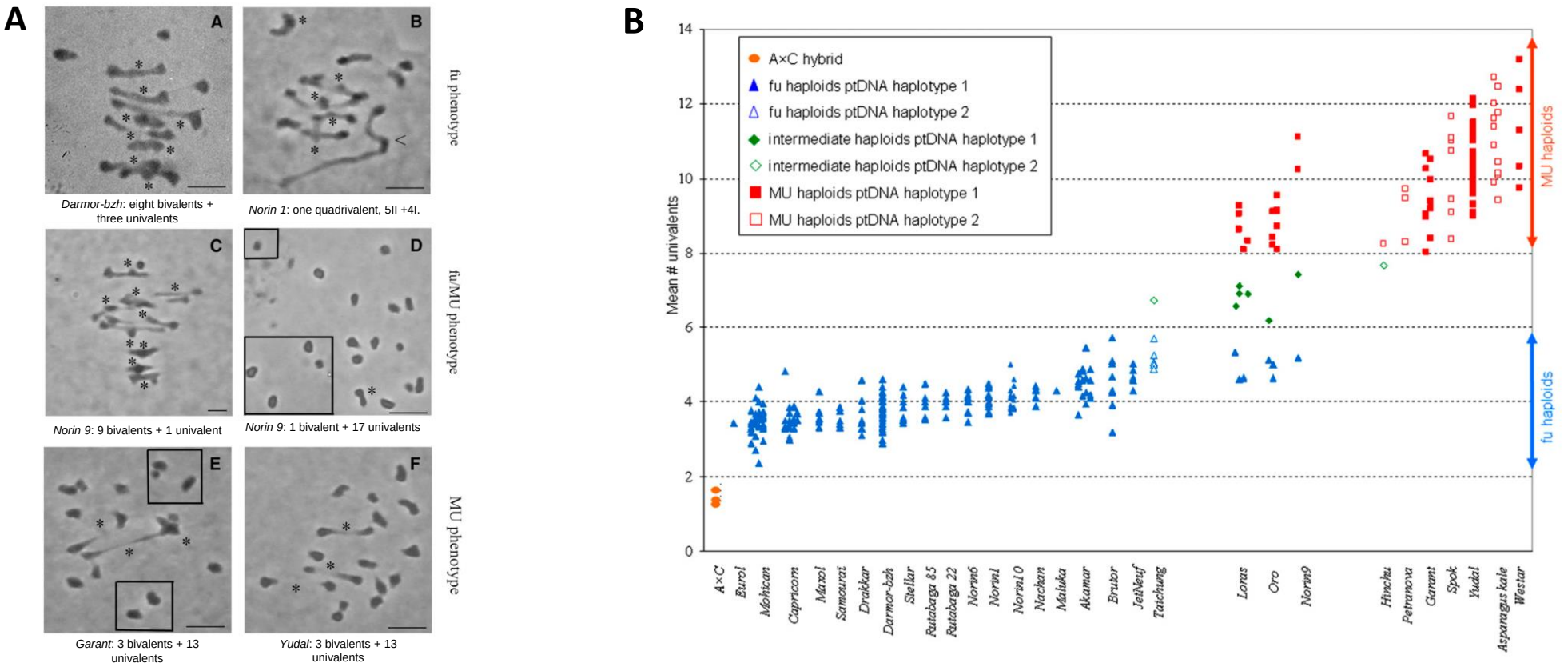


Figure 17 :Natural variation in CO frequencies in *B. napus* allohaploid [reproduced from (Cifuentes et al., 2010)]

(A) Nuclei of various *B. napus* varieties at metaphase I showing variation for the numbers of chromosomes that form bivalents. A-B: *Darmor*-like phenotype; C-D; mixture of *Darmor*-like and *Yudal*-like phenotypes produced from the same variety; E-F: *Yudal*-like phenotypes. The univalents located peripherally (out of the frame of these high magnification micrographs) are indicated within squares ([D] and [E]). Bivalents are indicated with an asterisk, and the quadrivalent in (B) is indicated with an arrowhead. Bars = 5 mm.

(B) Each symbol represents the mean number of univalents (calculated for 20 Pollen Mother Cells, i.e., PMCs) for every allohaploid plant isolated from the 29 *B. napus* accessions listed on the x axis and for five interspecific *B. oleracea* 3 *B. rapa* hybrids (noted AxC). Symbols with the same X-coordinate represent allohaploids isolated from the same plant. The clusters of consecutive X-coordinate samples represent three to four distinct plants sampled from the same population to account for its potential genetic heterogeneity (e.g., *Mohican*, *Capricorn*, etc.). Triangles represent allohaploids showing a high level of homoeologous recombination (*Darmor*-like allohaploids), diamonds represent allohaploids with an intermediate meiotic behavior, and squares represent allohaploids that showed a low level of homoeologous recombination (*Yudal*-like allohaploids).

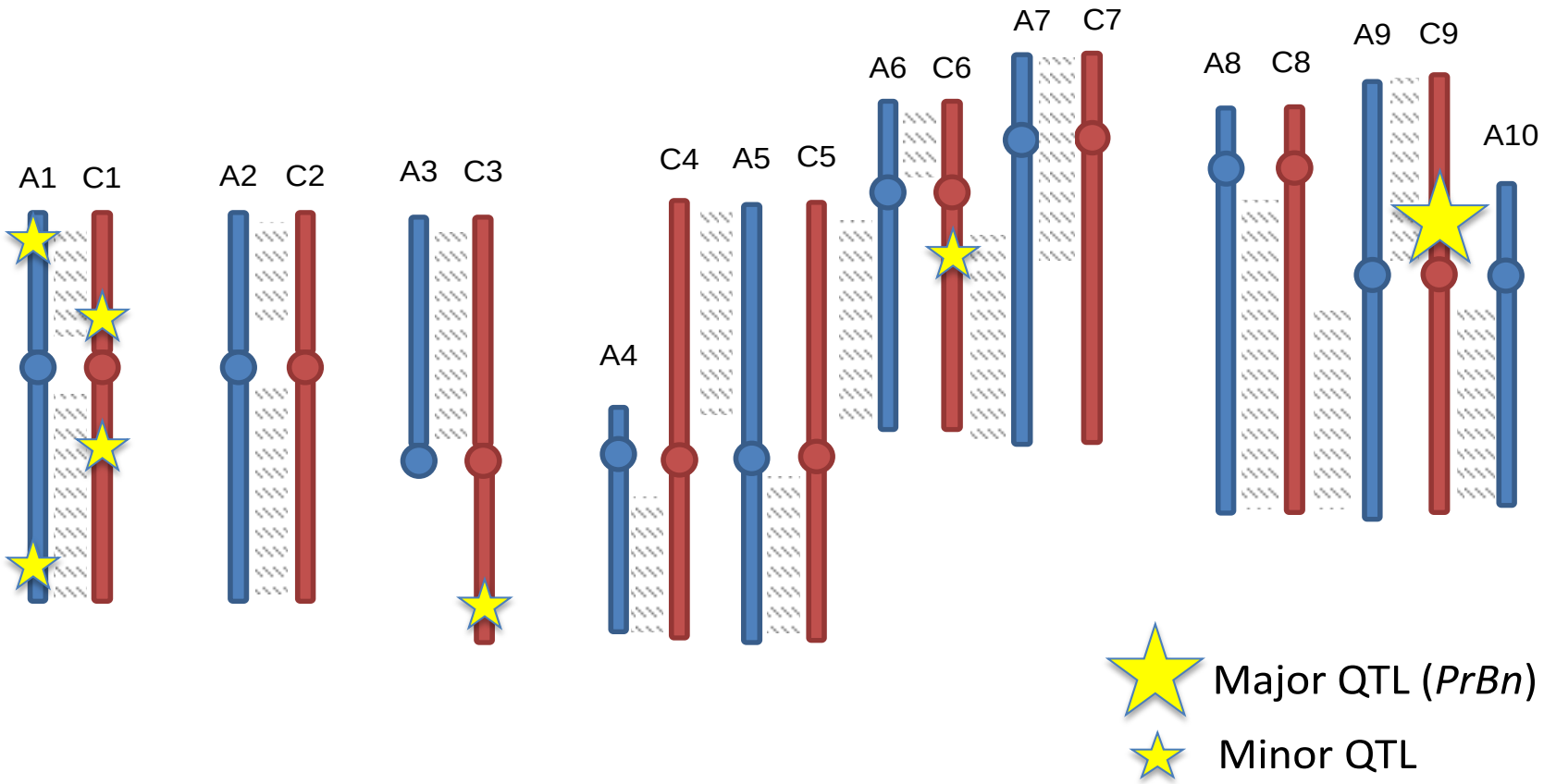


Figure 18: Mapping of *PrBn* and epistatic QTLs [adapted from (Liu et al., 2006)]

Representation of A (A_1-A_{10}; blue) and C (C_1-C_9; red) chromosomes of *B. napus* with the approximate position of *PrBn* (big yellow star) and 6 other epistatic QTLs for the control of CO frequencies between homoeologous chromosomes in allohaploids *B. napus*. Homoeologous regions between subgenomes are joined though hatchings.

Using a segregating population of allohaploids produced from *Darmor-bzh* x *Yudal* F1 hybrids, (Jenczewski et al., 2003) mapped a major locus (PrBn) on chromosome C_n09 that accounts for 40% of the explained variation. Four to six other additive or epistatic loci (Liu et al., 2006) were identified on the chromosomes A01, C01, C03 and C06 (Figure 18). Cifuentes et al. (2010) subsequently confirmed that segregation of two alleles at *PrBn* could adequately explain a large part of the variation in meiotic behavior found among *B. napus* allohaploids (Figure 17).

The presentation I have made so far might give the feeling that natural variation in CO frequencies between homologous chromosomes in euploids *B. napus* and natural variation in CO frequencies between homoeologous chromosomes in allohaploids *B. napus* are two distinct mechanisms. This conclusion would be abusive. Actually it is not known whether *PrBn* acts to suppress CO formation between homoeologues (with different efficiencies in *Darmor-bzh* compared to *Yudal*) or whether it plays a more general role in CO formation. The differences of CO frequencies observed in allohaploids and euploids might simply be two sides of the same coin; i.e. reflect a general increase of CO frequencies in *Darmor-bzh* compared to *Yudal* which is manifested between homologues in euploids and between homoeologues in allohaploids (because these are the least divergent partners in this context).

In that respect, *Brassica napus* would not necessarily be an exception within allopolyploids; there is indeed some indication that *Ph1* in wheat also affect CO formation between homologs in addition to limiting CO formation between homoeologs (Lukaszewski and Kopecký, 2010), and references therein).

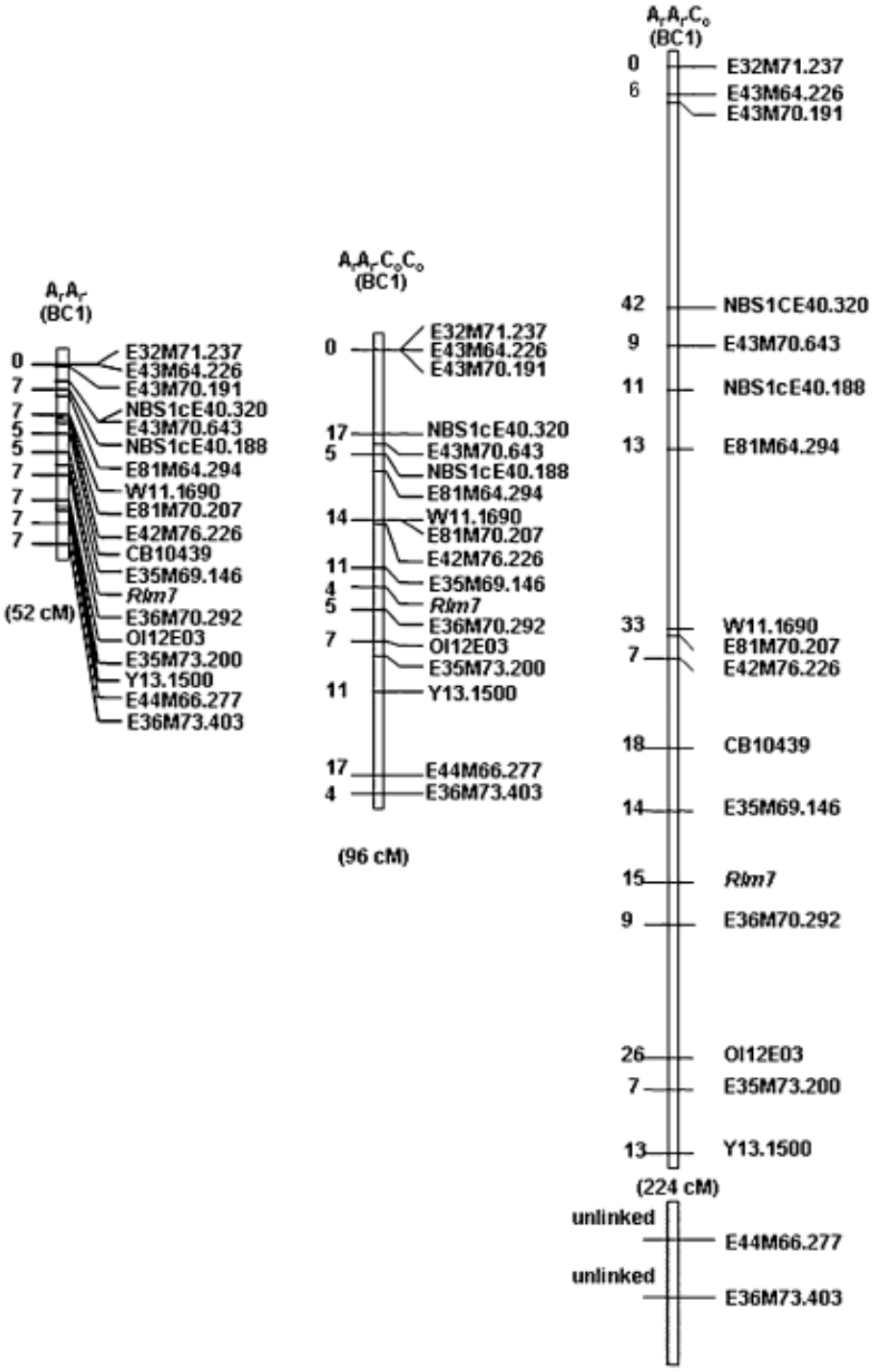


Figure 19: Genetic maps of the A7 Linkage Group in Progeny of the Diploid (ArAr), Allotriploid (ArArC), and Allotetraploid (ArArCC) Hybrids [reproduced from (Leflon et al., 2010)]

Genetic distances, indicated on the left of the linkage group, are expressed in cM and represent the distance between the marker and the annotated marker immediately above. The cumulative genetic size is indicated in brackets below each linkage group. Note that the 2 most distal markers at the bottom of the linkage group in the progenies of ArAr and ArArCC hybrids are not linked anymore in ArArC.

2.4 AAC triploid hybrids

Recent studies performed with allotriploid (AAC) *Brassica* hybrids gave further insights into the link between ploidy and CO frequencies (introduced previously, see p. 29). (Leflon et al., 2010) analysed meiosis of *Brassica* hybrids with the same genomic background but with different karyotypes. These authors found that CO frequencies on chromosome A07 increased in the progeny of allotriploid (ArArC) and allotetraploid (ArArCC) compared to the diploid (ArAr) hybrids; the highest CO rate (by far) being observed in the ArArC hybrid (Figure 19). Furthermore, the magnitude of the increase in AAC hybrids appears to be genotype dependent; triploids produced using *Darmor-bzh* (ArAdCd) made more COs than triploids produced using *Yudal (*ArAyCd) (Nicolas et al., 2009). More recently, it was shown that the number and the nature of the chromosomes that are left as a univalent modulate CO frequencies in *Brassica* triploids; interestingly addition of single chromosome C09, on which PrBn is located, is sufficient to boost CO frequencies (Suay et al., 2014).

Interestingly, and contrary to the anti-CO proteins (see p.15), at least some of the extra-COs observed in the triploids arise from the CO I pathway (dependent on ZMM proteins); (Leflon et al., 2010) observed an increase in the number of chiasmata marked by MLH1 during male meiosis (1,7 fold increase in the triploids compared to the diploid). The single increase of MLH1 foci is however insufficient to account for the almost 3-fold increase in genetic distances observed for female meiosis when comparing interval length between triploids and diploids (Leflon et al. 2010; Pelé et al., subm). Although (Suay et al., 2014) observed a drastic loss of interference in the triploids for almost all the genetic intervals they compared, they concluded that this could not result only from a massive increase in class II CO.

There are still very little insights into the mechanisms that drive the extra-CO formation in the triploids. It is interesting to note that the situation in the triploids echoes what is known about the control mechanisms that depend on the good progression of meiotic recombination (See paragraph 1.3 before, p17). In *C.elegans*, failure of a single chromosomes pair to synapse result in a compensatory increase of CO on the chromosomes that are correctly synapsed in the same cell (Carlton et al., 2006).

Chapter 3: Objectives of the PhD

The overall aim of my PhD was to better characterize the mechanisms that can be used to increase CO frequencies in an allopolyploids crops, *B. napus.* More specifically, I have addressed the three following research questions:

- First, what is the basis of natural variation for the control of CO frequencies*?* I have addressed this question within the frame of the *B. napus* allohaploids model. One of the objective of the group is to identify the causal polymorphism for *PrBn*, the main QTL explaining variation in CO frequencies between homoeologous chromosomes (see chapter 2, p.49). This task represents a long-term endeavour and exceeds by far the scope of this PhD. Before I joined the group, a RNA-seq analysis has been performed on meiotic tissue in two representative *B. napus* varieties (*Darmor-bzh* and *Yudal*), for two ploidy conditions (euploids; AACC or allohaploids; AC), to gain more insight on what govern differences in CO frequencies between varieties in both ploidy contexts. My personal contribution to this analysis has been to identify the main sources of variation in the meiotic transcriptome in this dataset and to characterize their respective contribution to the total variation. I will present the outcome of this analysis in the chapter 4 of this manuscript.
- Next, can we produce anti-recombinant plants in *Brassica* crop species by mutating one of the multiple pathway that limits CO frequencies in the model species *A. thaliana?* I have used a translational biology approach to assess whether the activity of FANCM, the first anti CO protein identified in plants (see chapter 1, p.15), was conserved in *Brassica* crops. My goal was to assess to what extent CO frequencies can be increased in *Brassica* crops defective for FANCM. I will present the outcome of this analysis in chapter 5.
- Last, is it possible to combine the increase in recombination frequencies that is observed in *Brassica* allotriploids (see chapter 2, p.56) by mutating an anti-CO protein? My objective was to produce a *Brassica* allotriploid mutants for FANCM and assess recombination frequencies in these plants. However, for reasons that I will develop at the end of chapter 5 (p.186), this axis of my project has been discontinued and no result will be presented in this manuscript.

Chapter 4 Deciphering the main source of variation for the meiotic transcriptome of *B. napus*

4.1 Introduction

Meiosis is the specialized cell division that is essential for gamete formation and sexual reproduction in eukaryotes. Although meiosis shares many features with mitosis, it also encompasses unique processes and distinct regulation mechanisms (Ohkura, 2015). Progression though meiosis requires expression and fine regulation of specific genes that are different from those needed in non-meiotic tissues. For examples, meiosis in the fission yeast *Schizosaccharomyces pombe* is accompanied by waves of gene expression along its progression (Mata et al., 2007) that are driven by a cascade of meiosis-specific transcription factors (Alves-Rodrigues et al., 2016). In plants, transcription factors specific to (male) meiocytes were also detected (Li et al., 2012; Dukowic-Schulze et al., 2014a; Flórez-Zapata et al., 2014) but their precise role during meiosis has still to be unraveled. The only exception (to the best of our knowledge) is the meiotic PHD-finger protein MMD1/DUET, which was recently shown to facilitate the progression of meiotic chromosome condensation in *Arabidopis thaliana* by promoting condensin gene expression (Wang et al., 2016a). AtMMD1/DUET is also required for proper expression of AtJAS and AtTDM1, two proteins involved in spindle organization during meiosis II and cell cycle transitions, respectively (Andreuzza et al., 2015). As expression of AtDUET and its target proteins is stage specific (Andreuzza et al., 2015), it is possible that transcription of meiotic genes is under stagewise control in plants as in yeast.

Overall, there is limited understanding of how gene expression is regulated during plant meiosis (Zhou and Pawlowski, 2014). To address this question, a growing number of global transcriptome analyses have been performed specifically on plant meiotic tissues (using either whole anthers or isolated meiocytes,) over the last decade [reviewed in (Zhou and Pawlowski, 2014; Dukowic-Schulze and Chen, 2014)]. These studies aimed at (i) identifying new meiotic candidates in complement to reverse and forward genetics (ii) gaining more insights into the regulatory pathways (e.g. transcription factors, chromatin remodelling genes, small RNAs and long non-coding RNAs) that control meiotic gene expression. Altogether, these studies revealed that a very large number of genes are transcribed during meiosis: ~60% of the genes annotated in Arabidopsis (Chen et al., 2010; Yang et al., 2011), ~50% of those in maize (Dukowic-Schulze et al., 2014b) and more than 40% of the gene models predicted in sunflower (Flórez-Zapata et

al., 2014). In addition to these annotated gene models, a wide range of unannotated features appeared to be transcribed in meiocytes of both Arabidopsis and maize (Dukowic-Schulze et al., 2014a). The meiotic transcriptomes also appeared distinct from the somatic transcriptomes in the same species, allowing for the identification of genes that are up-regulated or exclusively expressed in meiocytes. The meiotic transcriptomes of Arabidopsis, maize and sunflower were also compared with one another, and showed to vary widely (Dukowic-Schulze et al., 2014a; Flórez-Zapata et al., 2014). Nevertheless, a set of transcription factors up-regulated in meiocytes of both Arabidopsis and maize was identified (Dukowic-Schulze et al., 2014a).

Surprisingly, very few studies assessed the extent to which the meiotic transcriptome could vary between individuals within a single species. Primig et al. (2000) compared meiotic expression profiles of two yeast strains that display distinct kinetics and efficiencies of sporulation to define the "core" loci meiotically regulated in both strain. They observed that the core set only contained 60% of meiotically regulated genes in each strain and identified a negative regulator of meiosis, which is differentially expressed and may thus contribute to the different progression of meiosis between the two strains. Intraspecific variation of meiotic transcriptome can also be envisaged as linked with individual recombination rate variation. As noted above (see paragraph 1.5, p.25), deletion of the *Ph1* locus results in activation of transcription of functional cdk2-like copies on 5A and 5D (Al-Kaff et al., 2008) which may lead to increased Cdk2-like activity (Greer et al., 2012). Likewise, transcription of TaASY1 is 20-fold increased in the absence of *Ph1* while reduced transcription of TaASY1 (below WT level) resulted in crossover formation between homoeologous chromosomes at metaphase I (Boden et al., 2009).

As described above (see paragraph 2.3, p. 49), natural variation for meiotic recombination progression exists in allotetraploid *Brassica napus* (AACC; 2n=38). Likewise, natural variation for CO formation between homoeologous chromosomes was observed among allohaploid plants (AC; n=19) from this species (Cifuentes et al., 2010). To explain why the efficient sorting of homoeologous chromosomes in euploids is not paralleled by an almost complete suppression of CO formation between homoeologues in the corresponding allohaploids (Grandont et al., 2014), Jenczewski et al. (2003) hypothesized that the locus/loci responsible for chromosome sorting in *B. napus* could be haplo-insufficient, with different residual efficiencies associated with the different alleles of *PrBn*.

The objective of this study is to assess the extent to which the meiotic transcriptome varies between *B. napus* cv *Darmor-bzh* and cv *Yudal*, the two genotypes that were used to map *PrBn*

(variation for CO frequencies in allohaploids; Liu et al., (2006) and that showed different progression of homologous recombination in euploids (Grandont et al., 2014). For this, we examined and compared the transcriptome of isolated meiocytes through a mRNA-Seq experiment.

During the first step of the analysis, it became clear that we had to take into account a source of variation that had not been anticipated. Indeed, from the very beginning I stumbled upon the presence of homoeologous exchanges (HEs), both segregating within or fixed differentially between *Darmor-bzh* and *Yudal* (Chalhoub et al., 2014), that affected the meiotic transcriptome of the two varieties. I thus decided to further characterize this unexpected source of variation before assessing the relative contribution of the other factors.

The result of the transcriptomic analysis will therefore be presented in two steps: (i) I will first characterize the impact of HEs on gene expression and show that how the presence of HEs can be used to further study the link between gene copy number and expression. For this analysis, I only used part of the data generated from euploid *Darmor-bzh* and *Yudal*. This work is presented as a manuscript, currently under review, for which I am first co-author. (ii) I will then present the main results of the differential expression analysis performed on the full dataset to assess the relative contributions of the other sources of variation. This part of the analysis is comparatively less advanced; as a consequence, these results are presented in a way that deviates from a research paper manuscript.

4.2 Manuscript: Homoeologous exchanges drive extensive dosage dependent changes in gene expression and influence allopolyploid genome evolution

This manuscript corresponds to the version that is currenlty under review at Molecular Biology and Evolution.

Article (Discoveries)

Homoeologous exchanges drive extensive dosage dependent changes in gene expression and influence allopolyploid genome evolution

Andrew Lloyd[1,a,b], Aurélien Blary[1,b], Delphine Charif [b], Catherine Charpentier[b], Joseph Tran[b,c,d], Sandrine Balzergue[c,d,e], Etienne Delannoy[c,d], Guillem Rigaill[c,d] & Eric Jenczewski[b]

[1]Joint first author
[a]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
[b]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, 78000 Versailles Cedex, France.
[c] Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, Bâtiment 630, 91405 Orsay, France
[d] Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405, Orsay, France.
[e] IRHS, INRA, AGROCAMPUS-Ouest, Université d'Angers, SFR 4207 QUASAV, 42 rue Georges Morel, 49071 Beaucouzé cedex, France

Corresponding authors
Andrew Lloyd: andrewhmlloyd@gmail.com
Eric Jenczewski: eric.jenczewski@versailles.inra.fr

Keywords
gene expression, transcriptome, polyploid, allopolyploid, homoeologous, homoeologous exchange, crop, structural variation, gene dosage

Abstract

Structural variation is an important substrate for natural selection. In allopolyploid species, homoeologous exchanges (HEs; i.e. genomic exchanges between the constituent subgenomes) are a significant source of structural variants, which lead to large changes in gene content and copy number. We show here that HEs contribute to gene expression variation in *Brassica napus*, a representative allotetraploid crop and that the HEs result in expression changes proportional to the change in gene copy number. HEs, therefore, contribute to major differences in gene expression between and within varieties, defining large clusters of genes with consistently increased or decreased expression. When homoeologous gene pairs have unbalanced transcriptional contributions prior to the HE, duplication of one copy does not accurately compensate for loss of the other and combined homoeologue expression also changes. This trend is less pronounced for older HEs, suggesting that transcriptional compensation between homoeologues or selection against some HEs (or HE segments) has buffered total homoeologue expression over time Thus, the interplay between new structural variants and their resulting impacts on gene expression, influence allopolyploid genome evolution.

Introduction

Structural variation, i.e. large scale genomic alteration, is a major source of genetic diversity both within (Pezer et al., 2015) and between species (Perry et al., 2006). Structural variants have been shown to be responsible for a broad range of phenotypes, including severe genetic disorders in humans (Weischenfeldt et al., 2013b) and many developmental changes in plants (Saxena et al., 2014; Zhang et al., 2015). The mechanisms by which structural variants are formed and convey phenotypes are broad (Carvalho and Lupski, 2016; Weischenfeldt et al., 2013b), but often involve the deletion or duplication of dosage-sensitive genes through recombination-related processes (Bai et al., 2016; Carvalho and Lupski, 2016). The induced effect on gene expression is not necessarily limited to the genes within the rearranged region, but may result in wide-spread differential gene expression (Henrichsen et al., 2009; Guryev et al., 2008).

In plants, polyploidy or whole genome duplication (WGD) is a prominent force driving structural variation. This is especially true in allopolyploids that combine two or more differentiated genomes, referred to as homoeologues. Over evolutionary time scales (millions of years), duplicated genes tend to return to a single copy following polyploidy, with a decreasing rate of gene loss through time (Lloyd, 2014; Li et al., 2016). Current evidence

indicates that gene loss starts soon after allopolyploid formation, although few genes are affected over the first hundreds of generations (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2014; Chalhoub et al., 2014). In these early stages, a mechanistically distinct form of gene loss (Woodhouse et al., 2010; Gaeta and Chris Pires, 2010) results from segregation of large homoeologous exchanges (HEs). While HEs initially arise through crossover-driven reciprocal exchanges between homoeologues (Nicolas et al., 2007; Gaeta and Chris Pires, 2010), only one exchanged region usually becomes fixed within the population. This results in the replacement of one chromosomal region (which is lost) with a duplicate of the corresponding homoeologous region. For the sake of simplicity, even if the term has some drawbacks, we will refer to these fixed structural variants as "HEs" as per Chalhoub et al. (Chalhoub et al., 2014).

HEs are commonplace in very recent allopolyploid genomes (Chester et al., 2012b; Gaeta and Chris Pires, 2010) as well as in slightly older allopolyploid species: *Brassica napus* (Chalhoub et al., 2014), *Gossypium hirsutum* (Li et al., 2015), *Coffea Arabica* (Lashermes et al., 2014), *Arachis hypogaeae* (Bertioli et al., 2016). Despite this apparent pervasiveness, much remains to be learned about the biological attributes of HEs. For example, little is known about the extent to which, and the means by which, HEs contribute to the loss of dispensable genes in allopolyploid species or drive differential gene expression between genotypes within these species. A series of papers in the second half of the 2000s attempted to address this latter question (Tate et al., 2006; Gaeta et al., 2007, 2009; Koh et al., 2010), however the technologies available at the time to survey gene expression had inherent limitations. As a result, these assessments dealt either with a handful of qualitative markers (e.g. presence/absence of cDNA-AFLP markers; (Tate et al., 2006; Gaeta et al., 2007; Koh et al., 2010) or were unable to distinguish between homoeologous transcripts (Gaeta et al., 2009). Thus, while more re-arranged genomes were correlated with more divergent transcriptomes, the scope of the results obtained and their interpretation are limited and the impact of HEs on gene expression remains largely an open question.

In this study we used a tissue-specific mRNA-seq dataset to measure the consequences of homoeologous exchanges on gene expression. To provide a sound basis for this analysis, we first re-evaluated the HE landscape in two representative *B. napus* cultivars, following the seminal work of Chalhoub et al. (Chalhoub et al., 2014). Thanks to the mRNA-seq analysis, we were able to quantify and compare transcript levels while distinguishing the contribution of the two homoeologues prior to, and after, the HEs had occurred. The use of a single tissue, in which a large percentage of genes are expressed (Chen et al., 2010), enabled us to maximise genome

coverage, while eliminating potential noise introduced due to different contributions of homoeologues to the transcriptome in different tissues (Adams et al., 2003; Chalhoub et al., 2014). The substantial inter-individual variations in gene content and expression that we observe provide new insights into the impact of HEs on the evolution of allopolyploid genomes.

Results

Homoeologous Exchanges are distally located, affecting regions of high gene density and high recombination frequency in *Brassica napus*

To ensure accurate expression analysis, we first determined whether the HEs in our plant lines were identical to those reported in Chalhoub et al (Chalhoub et al., 2014), validating their presence in *Brassica napus* cultivars *Darmor-bzh* and *Yudal*. For this, we used single-nucleotide polymorphism (SNP) genotyping (Brassica 15k array) and direct Sanger sequencing of consensus PCR amplicons to test for the expected loss of one of the two homoeologous copies associated with every HE (Fig. 1A; Fig. S1). This approach confirmed the loss of one homoeologue for 27 of the 30 previously identified HEs (15/17 *Darmor-bzh*; 12/13 *Yudal*; Table S1-2). For the three remaining events, both A and C genome copies were shown to be present (Fig. S1), suggesting that the events may be absent from our lines, or involve more complex rearrangements than simple HEs; thus they were precluded from further analyses.

We next examined whether closely located HEs were indeed interspersed with non-exchanged areas or should rather be grouped to form a single, longer HE. In most cases closely spaced HEs were indeed interrupted by non-exchanged areas (Fig.1A; Fig. S1). For the series of HEs initially described along A_n1 and C_n1 we detected additional $A_n1^- \leftarrow C_n1^+$ (i.e. A_n1 replaced by C_n1) events in *Yudal*, such that overall *Yudal* HEs ($A_n1^- \leftarrow C_n1^+$) mirrored the patchwork of interspersed HEs ($A_n1^+ \rightarrow C_n1^-$) and non-exchanged regions observed in *Darmor-bzh* (Fig. S1). We think it is improbable that this symmetrical patchwork pattern could have arisen by chance independently. Rather, we expect that these represent two contiguous HEs (one in *Darmor-bzh*, one in *Yudal*) and that their patchwork appearance reflects problems with the underlying pseudomolecule assembly rather than the true chromosomal order of the HEs. We therefore considered these events contiguous HEs for further analyses (Fig. 1A).

All HEs were located within the most distal third of chromosome arms (Fig. 1B), many of them being even closer to the telomeres; this notwithstanding, only 6 events (out of 27) extended to chromosome ends (Fig. 1A). This distal distribution of HEs reflected local homologous

recombination rates; i.e. HE frequencies increase with homologous crossover rates along the average chromosome arm (Fig. 1B; $R^2 = 0.92$, p = 3.7E-6). HEs are also located in regions of high gene density (Fig. 1B). We estimated that the HEs fixed either in *Darmor-bzh* or in *Yudal* together encompass a few thousand (> 3,500) gene models (Table S3); this is an under-estimate, however, because for many HEs, the duplicated regions are absent (or partly absent) from the *B. napus* assembly (Chalhoub et al., 2014) making it impossible to assess the number of gene models within these regions.

Homoeologous Exchanges generate clusters of differentially expressed genes

Following an HE, gene loss is usually accompanied by replacement with its homoeologue. This results in the establishment of two identical gene copies (e.g. AC → AA) that, unlike many copy number variants, segregate independently. In the classical sense, two independently segregating loci constitute two genes; however, for HEs it seems more relevant to consider a gene the duplicated loci that contribute to the expression of a unique mRNA (Fig. 2A). This is biologically relevant as the same mRNA produced from two independent loci will have the same phenotypic consequences, and also methodologically relevant, as while we can distinguish homoeologues (i.e. A vs C), it is impossible to distinguish sequencing reads that originate from two identical, but independently segregating loci (i.e. A vs A or C vs C).

Based on this premise, we first determined whether the HEs we had confirmed, generate divergent gene expression profiles. To do this we compared the expression profiles of *Darmor-bzh* and *Yudal* in regions outside HEs (representing the baseline divergence between the two varieties) and within HEs. Our results not only confirmed the expectation that regions lost in *Yudal* were enriched in down regulated genes but also demonstrated that the corresponding duplicated regions in *Yudal* were enriched in upregulated genes compared to *Darmor-bzh* (Table S4). This holds true for three of the HE-driven duplicated regions in *Darmor-bzh*, which were enriched in up-regulated genes, however it was not possible to evaluate the equivalent regions lost in *Darmor-bzh*, as they are not present in the reference genome assembly (Chalhoub et al., 2014).

We then tested whether HE expression profiles are sufficiently different from genome average to be identified without any prior indication of their position. Given that series of adjacent genes are lost or duplicated as a consequence of HEs, we looked for clusters of genes with a consistent direction of transcriptional change. In accordance with previous results, segmentation of gene expression (Fig. 2B) identified all confirmed HEs in *Yudal*; all lost regions were detected as under-expressed segments, and 6 out of the 13 concurrently duplicated regions were detected

as over-expressed segments compared to *Darmor-bzh* (Tables S5-7). This approach also detected two clusters of genes displaying similar patterns, but that did not overlap with known HEs. The validation procedure described above (SNP/PCR; Tables S2 and S8) was applied, confirming that the corresponding regions were lost in *Yudal* (Fig. S1). It is thus likely that these two clusters of genes correspond to additional HEs. By contrast, none of the 13 *Darmor-bzh* HEs were identified by the segmentation analyses (Tables S5-7); this is likely due to the partial assembly of these regions of the reference genome (Chalhoub et al., 2014), which reduced our statistical power to detect these events *de novo*.

Finally, we observed that genes in HEs had a disproportionate effect on the total transcriptome. While the affected genes represent less than 4% of total gene number, they represent a larger percentage of those with highest (absolute) fold-change between cultivars: i.e. 22% of the top 1% (fold-change > 252; χ^2, p = 9.5E-100) and 19% of the top 5% (fold-change > 8.9; χ^2, p = 0). Although we are using a single cell type, these results are a good representation of the genome-wide effects of HEs as 63% of all genes are transcribed in our data set (68% and 45% of these being covered by >10 and >100 reads per sample respectively).

Segregating HEs drive massive gene expression changes within a variety

We next investigated whether any equivalent regions existed between biological replicates within a variety. For the three *Darmor-bzh* biological replicates, we identified a pattern of expression that was evocative of HEs previously identified, in a single chromosomal region at the top of A_n1-C_n1 (Figure 2C). PCR confirmed the physical loss of one (ACCC) or two (CCCC) copies of the A genome in this region (Fig. S2). No equivalent regions were identified in *Yudal*. These results indicated that a newly-formed HE was segregating among *Darmor-bzh* biological replicates. Contrary to the previously observed HEs, fixed either in *Darmor-bzh* or *Yudal*, this segregating event encompassed a very large region (4.4 Mb or 1470 genes). We compared the gene content between the two exchanged homoeologous regions using the synteny tool within the Genoscope *Brassica napus* genome browser and identified a total of 43 gene models that are specific to the A region (Fig. S3); as these genes have no homoeologue, their loss cannot be compensated in the CCCC genotype. More broadly the HE had a very large effect on the total transcriptome, with affected genes representing the majority of those with highest (absolute) fold-change between the AACC and CCCC genotypes; 94% of the top 1% (fold-change > 34; χ^2, p = 0) and 47% of the top 5% (fold-change > 1.7; χ^2, p = 0). This segregating event offered a unique opportunity to evaluate the extent to which variation in gene copy number correlates with gene expression change.

A vast majority of genes show additive expression when duplicated in the newly formed HE

Overall we observed that the level of expression of a gene in the newly formed HE was directly proportional to the number of copies of that gene, with the expression ratio being very close to, or equal to, the ratio of gene copy numbers between *Darmor-bzh* biological replicates (Fig. S4-5): e.g. A-copy gene expression decreased twofold, while C-copy gene expression increased 1.5-fold between AACC and ACCC genotypes. Only 69 genes (out of 1470; 4.5%) deviated significantly from this general trend. These outliers were enriched in homoeologous pairs (16 pairs; χ^2, p = 5.5E-21), indicating that homoeologues are likely to respond similarly to gene dosage variation. Remarkably, of the 69 outliers, 42 (60.8%) showed decreased expression when copy number increased. Despite these outliers, for the vast majority of genes (95.5%) copy-specific gene expression was in strict concordance with gene copy number immediately following an HE (Fig. 2D). As a consequence, differences in the summed expression of homoeologues (hereafter, $Total_{(A+C)}$ expression), depended on the relative contributions of the two copies prior to the HE (estimated from the AACC genotype). This represented a continuum where duplication of a dominantly expressed homoeologue led to increased $Total_{(A+C)}$ expression and duplication of a lesser expressed homoeologue led to reduced $Total_{(A+C)}$ expression (Fig. 2D). For this reason, almost half (43%) of the homoeologous pairs affected by the newly formed HE had significantly altered $Total_{(A+C)}$ expression (Fig. 2D). Conversely, for 57%, $Total_{(A+C)}$ expression remained unchanged in the CCCC and ACCC genotypes. This latter group corresponded to genes where the A and C copies contributed equally to $Total_{(A+C)}$ expression in the AACC genotype (Fig. 2D, Homoeologue Bias ~ 0.5).

Most genes within older fixed HEs also show additive expression, but additional factors contribute

To gain insights into longer-term effects of HEs on gene expression we analysed the impact on gene expression of the older HEs fixed in *Yudal* (Chalhoub et al., 2014). Unlike the case above, this analysis was constrained by the lack of a direct pre-HE reference genotype for comparison. Instead we used the expression pattern observed in *Darmor-bzh* as a proxy for the pre-HE state in *Yudal* (Fig. S6). This necessary approach potentially introduced additional layers of transcriptional variation and also reduced the number of HEs amenable to analysis (i.e. *Yudal* HEs that overlap with *Darmor-bzh* HEs cannot be used). In spite of this, we still observed that expression of genes within the fixed HEs in *Yudal* was essentially dosage dependant; most genes duplicated by HEs in *Yudal* showed an almost 2-fold increase in expression compared to

that of their single copy homolog in *Darmor-bzh* (Fig. S7). However, the absolute dose difference did not appear to be the only determinant of gene expression in *Yudal* HEs (Fig. 2D, Fig. S8).

To confirm additional influences on expression for genes within fixed *Yudal* HEs, we compared globally, the concordance in *Darmor-bzh* and *Yudal* per-copy expression levels for genes within HEs and for genes outside HEs. If gene expression is purely additive, then these two distributions should be similar. This approach also enabled us to isolate the effects on gene expression attributable to HEs from those due to inter-varietal variation. A two-sample Kolmogorov-Smirnov test verified that the two distributions (concordance in per-copy *Darmor-bzh* and *Yudal* expression levels, inside and outside HEs) differed significantly (Fig. S9, p = 3.2E-4), confirming divergent transcriptional output for genes within HEs. While this test demonstrated that the distributions differed, it provided little insight into why. Further analyses, however, shed some light on the drivers of this divergent transcriptional outcome.

Fixed HEs have gene content and transcriptional outputs consistent with transcriptional compensation and/or selection against HEs (or HE segments) over time

Firstly, we observed that within fixed *Yudal* HEs there were fewer instances where the lesser-expressed homoeologue was retained and more instances where the dominantly-expressed homoeologue was retained than expected, given the distribution of A:C ratios genome wide (Tables 1 & S10). In spite of this, when lesser expressed copies were retained in HEs, they tended to show the expected behavior, with reduced $Total_{(A+C)}$ expression (Table 1). Similarly, duplication of the dominantly expressed homoeologue mainly lead to increased $Total_{(A+C)}$ expression (Table 1). Surprisingly, duplication of the dominantly expressed homoeologue also led to unexpectedly frequent instances of decreased $Total_{(A+C)}$ expression (Tables 1 & S10).

We next compared the observed $Total_{(A+C)}$ expression to that expected based on predicted pre-HE levels. Here we restricted our comparisons to those homoeologues with a strong bias (AC log2.ratio < -2 or > 2) as these genes are expected to have their $Total_{(A+C)}$ expression most affected by an HE. While there was no divergence from expectation for genes within the newly-formed HE (Fig. 2E), for fixed *Yudal* HEs we observed that duplication of the lesser expressed copy resulted in higher $Total_{(A+C)}$ expression than expected (Fig. 2F). A similar trend held for genes outside HEs, but the magnitude of this change was lower than that for within HEs (Fig. 2F). These results are concordant both with transcriptional adjustment to re-establish prior expression levels and/or selection against poorly expressed genes in HEs. Consistent with this hypothesis, the proportion of homoeologous pairs within the fixed HEs that showed differential

$Total_{(A+C)}$ expression between *Darmor-bzh* and *Yudal* was not different from that observed outside HEs.

Discussion

The presence of a few homoeologous exchanges has long been reported in *B.napus* (Sharpe et al., 1995; Butruille et al., 1999; Lombard and Delourme, 2001; Piquemal et al., 2005; Udall et al., 2005; Howell et al., 2008; Osborn et al., 2003) although their origin and biological impact has remained unclear. In this study we assess the immediate and long term impact of HEs on gene expression and provide new insights into the transcriptomic consequences of HEs as well as their origin and evolutionary fate.

Our first observation, during routine confirmation of HEs, was that the plants used in this study differed in their HE content by ~15% from that published by Chalhoub et al (Chalhoub et al., 2014) even though all plants originated from the same highly inbred varieties. These discrepancies may not only stem from the different approaches used to detect HEs, but probably also reflect a biological reality. This point is perhaps more strikingly made through the fortuitous detection of a newly-formed HE segregating in *Darmor-bzh* (Fig. 2C), showing that even siblings may differ in their HE content.

The continuous emergence of novel HEs, which has been previously observed (Sharpe et al., 1995; Udall et al., 2005), reflects ongoing, rare recombination between homoeologous chromosomes in established euploid *B. napus* (Grandont et al., 2014). This also explains why very few HEs are shared between *Darmor-bzh* and *Yudal*, as these events have most likely accumulated gradually, from the occurrence of independent, sporadic COs between homoeologues. Given the variable nature of HE content between (and within) *B. napus* varieties, determining the particular events present within the lines studied (as we have done here) is important to lay the foundation for comprehensive RNAseq analysis in this species.

A second observation is that the newly formed HE had specific features that contrast sharply with those of the older HEs. Whereas older HEs are relatively small, occasionally interspersed with non-exchanged areas and usually sub-terminal, the newly formed HE is large, un-fragmented, and extends to the end of the chromosome; in this respect it is similar to the majority of newly-formed HEs generated during meiosis of *B. napus* allohaploids (Nicolas et al., 2012). As HEs are located in regions where recombination frequency is the highest (Fig.

1B), it is tempting to speculate that further crossovers ultimately re-shape initially large HEs into several smaller events. Under the hypothesis that large HEs, which occur in gene dense regions (Fig. 1B), may be slightly deleterious and thus selected against (Szadkowski et al., 2011; Gaeta and Chris Pires, 2010), positive selection of even rare recombination events would result in the preferential break up of large HEs and thus determine which genes ultimately become duplicated or lost within the genome.

Even if initially large HEs are refined to smaller segments through time, we show here that they are still major contributors to gene loss, accounting for a >3,500 gene content difference between two representative *B. napus* genotypes (Table S3). Given that this gene loss does not impact the viability of these two allopolyploid *B. napus* genotypes, HEs can be considered major contributors to the loss of dispensable genes i.e. the acknowledged set of genes that is only present in some but not all individuals of a species (Albalat and Cañestro, 2016). Part of this "dispensability" can be attributed to the buffering effect of the presence of homoeologous copies within the same genome (Lim et al., 2008; Xiong et al., 2011). This is exemplified by our observation that the ($A_n1^- \leftarrow C_n1^+$) events in *Yudal* mirror the ($A_n1^+ \rightarrow C_n1^-$) events in *Darmor-bzh* (Figure S1) indicating that either of the two copies can be lost and functionally replaced by its homoeologue in these regions.

However, such cases should not be used to conclude that all homoeologous copies are interchangeable. Homoeologous regions may differ locally in gene content (Feuillet et al., 2001; Griffiths et al., 2006); for example we estimated for the newly formed HE reported here, that 43 A-genome specific genes were lost in the CCCC genotype (Fig. S3). Homoeologues may also differ in their coordination with cytoplasmic organellar genomes (Gong et al., 2014; Sehrish et al., 2015), which may in part explain the preferential loss of C copies (Fig. 1A; 22 events out of 27; χ^2 test; p=0.009) observed in *B. napus* (Nicolas et al., 2012; Chalhoub et al., 2014). These quantitative and qualitative differences in gene content between homoeologues ensure that HEs will generate new variation on which selection can act, the result of which will ultimately influence which HEs are retained and which are lost or re-shaped.

In addition to being major contributors to differences in gene content, HEs have profound impacts on gene expression. Following homoeologous exchanges, changes in gene copy number result in proportional changes in gene expression (Fig. S4), with the ratio of mRNA abundance being almost equal to that of gene copy number. This is reminiscent of the impact of other types of copy number variation on gene expression (Zhang and Oliver, 2007). HEs therefore result in clusters of genes with divergent expression profiles, not only between

separate accessions (Fig. 2B) but also between individuals within an accession (Fig. 2C). These clusters of genes have a profound impact on the total transcriptome in *B. napus* with a large fraction of the genes with high fold-change between (and within) varieties found in regions of the genome affected by an HE.

The anticipated consequences of HE-driven changes in gene expression, however, depend on the unit of transcription that is biologically relevant. If homoeologues have divergent function e.g. because of extensive divergence in alternative splicing between homoeologues (Zhou et al., 2011; Chalhoub et al., 2014) or tissue specific expression patterns (Adams et al. 2003; Chalhoub et al., 2014), then their transcription should be considered independently and HEs are likely to contribute to phenotypic variation (Pires et al. 2004; Chalhoub et al., 2014). If homoeologues are functionally redundant however, $Total_{(A+C)}$ expression is the relevant measure and no phenotypic effect is expected if the loss of one copy is buffered by the duplication of its homoeologue. While it might be expected that such buffering occurs (Xiong et al., 2011), we observed that in just under half of the cases (43%) duplication of one copy did not compensate for the loss of the other immediately following an HE i.e. $Total_{(A+C)}$ homoeologue expression was different in the three genotypes (e.g AACC ≠ ACCC or CCCC; Fig. 2D).

There are at least two instances where duplication of one homoeologue does not restore the expression contributed by the lost copy. This happens either when **i)** expression is inversely proportional to gene copy number; this occurs for only a minority of genes (3%), and **ii)** homoeologues are differentially expressed prior to the HE (Fig. 2D). In this second case, duplication of the dominantly expressed homoeologue results in increased $Total_{(A+C)}$ expression, while duplication of a lowly-expressed copy results in reduced $Total_{(A+C)}$ expression.

These trends are reminiscent of two observations made in studies of gene expression in newly formed (neo-)polyploids; a) that genes whose expression differs from the mid-parent value (i.e. average expression of the two parents) are enriched in those that are differentially expressed between the progenitors (Wang et al., 2006; Gaeta et al., 2009) and b) that such genes primarily adopt the expression level of one of the two parents (reviewed in Yoo et al., 2014). Both of these observations may be explained by the accumulation of HEs, which occur at high rates in neo-polyploids (Gaeta et al., 2007; Henry et al., 2014). If differences in gene expression

between polyploid progenitors are mediated by cis-regulatory elements (Chaudhary et al., 2009; Shi et al., 2012; Combes et al., 2013), then these differentially expressed genes will likely show a homoeologue bias in neo-polyploids. In these instances, duplication of the dominant homoeologue will result in expression approaching that of the parent with high expression, while duplication of a lowly expressed homoeologue will result in result in expression approaching that of the parent with weak expression. This also explains why the set of genes whose expression differs from the mid-parent value is different in each re-synthesized line (Gaeta et al., 2009), as each independent line will accumulate a unique set of HEs.

The immediate effects of HEs on expression are clearly strong, with >40% of homoeologues in our study showing immediately altered $Total_{(A+C)}$ expression which was mostly explained by a bias in homoeologue expression prior to the HE. Given that the homoeologue bias is known to differ between tissues (Adams et al., 2003; Chalhoub et al., 2014), our use of a single cell type was likely important in providing a clear snapshot of the full diversity of A:C expression ratios and the corresponding effects of HEs on expression. The use of a single cell type also highlighted the full potential phenotypic impact of HEs as phenotypes are derived from expression at the cellular level, rather than the average expression of a complex tissue.

Given that such a large percentage of homoeologues showed altered $Total_{(A+C)}$ expression, and that transcription levels for some genes need to be tightly controlled to ensure fitness, HEs (or HE segments) that induce detrimental expression levels should be selected against. In this connection, we observed a body of corroborating evidence supporting the hypothesis that subsequent changes have occurred within older fixed HEs to mitigate the transcriptional burden initially introduced. We observed a reduction in the proportion of genes for which the duplicated copy (i.e. A genome copy) was the lesser expressed homoeologue, consistent with selection against HEs or HE segments containing lowly expressed genes. This selection is likely to be particularly strong for genes that show an inverse relationship between gene expression and gene copy number; the expression of these genes could also explain why duplication of the dominantly expressed homoeologue contributed to some cases of decreased $Total_{(A+C)}$ expression. Further, when comparing old and new HEs, we observed that $Total_{(A+C)}$ expression in AAAA (or CCCC) more closely resembles that seen in AACC for the older HEs (Fig. 2D). In addition, expression of lowly expressed homoeologues is higher than expected in the older HEs (Fig. 2E). These last two findings are consistent both with the aforementioned selection against lowly expressed genes, and/or transcriptional compensation to bring total expression of the homoeologous pair closer to that seen prior to the HE (e.g. Pala et al., 2008; Hose et al.,

2015). Subsequent evidence is required to ascertain which (if either) of these two processes contributes most to alleviation of the initial gene-dosage effects imposed by HEs.

Conclusion

Taken together our results suggest a nuanced model for the dynamics of structural variation and gene expression following HEs. Homoeologous exchanges are likely large initial events (Nicolas et al., 2012; Gaeta and Chris Pires, 2010) that affect a large number of genes. For most homoeologous genes, the HE will be tolerated with little to no change in $Total_{(A+C)}$ expression. For others however, particularly dosage sensitive genes with a strong A/C bias, selection will favor re-establishment of normal expression level. This may occur as a consequence of ongoing recombination which can locally restore the pre-HE state so long as the HE is not fixed, or through transcriptional compensation. This situation is probably not limited to *B. napus* but likely applies to all allopolyploid species with demonstrated (e.g. *G. hirsutum* (Li et al., 2015); *C. Arabica* (Lashermes et al., 2014); *A. hypogaeae* (Bertioli et al., 2016)) or as-yet unidentified HEs. The complex interplay between structural variation, and the resulting consequences on gene content and expression, therefore probably shapes the evolution of most allopolyploid genomes.

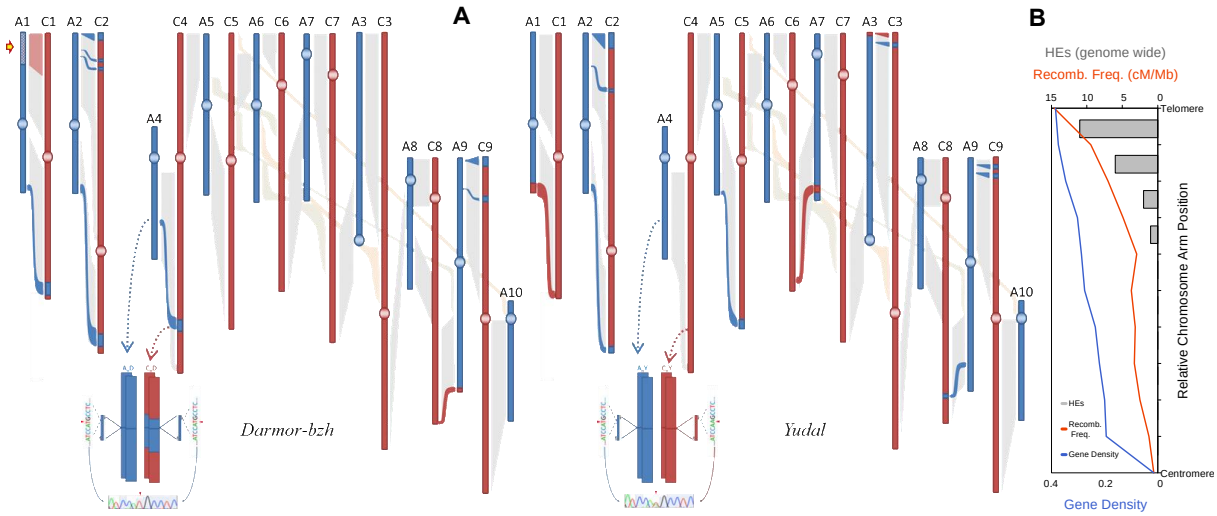


Figure 1.

A) Overview of Homoeologous Exchanges (HEs) in *Darmor-bzh* and *Yudal*. Homoeologous regions between A (A_1-A_{10}; blue) and C chromosomes (C_1-C_9; red) are joined though hatchings. Superimposed on this background are thick coloured lines that indicate the position of fixed HEs (blue: A duplicated/C lost; red: A lost/C duplicated). The newly-formed HE is marked by an arrow (top of A_1). Diagrams at the bottom of chromosomes A_4 and C_4 show the PCR amplicons and part of their sequences alongside the chromosomes; the corresponding chromatogram profiles (bottom) are centered on a SNP discriminating the two sub-genomes that we used to test for the presence or absence of HE. **B) HEs occur distally in regions with high recombination frequency and high gene density.** Gene density (% genic sequence, blue) and recombination frequency (cM/Mb, red) along the average chromosome arm containing an HE in *B. napus*. The cumulative number of HEs in each bin (genome wide) is given by the histogram.

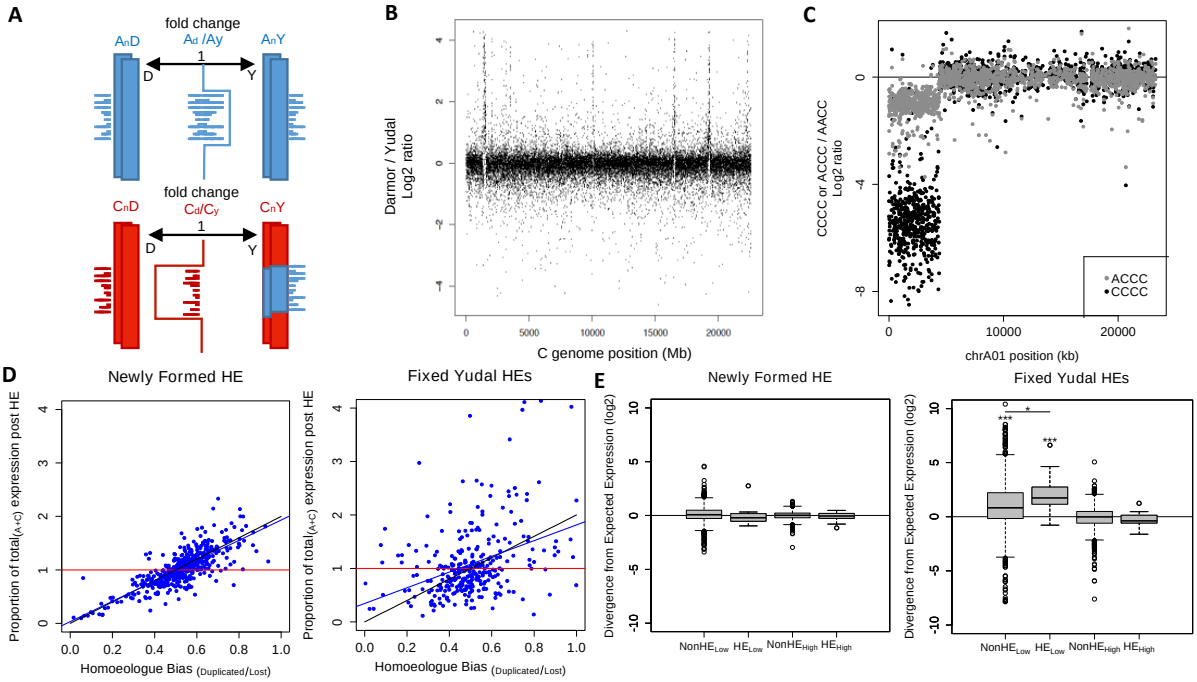


Figure 2. A) Theoretical model: The dashes alongside the chromosomes represent the sequenced mRNA reads from the genes located in the corresponding A (blue) and C (red) regions of an HE in *Yudal*. The sum of those reads which map on the reference genome appears in between the chromosomes, as well as the mean *Darmor-bzh* / *Yudal* expression ratio. This ratio is centred on 1 except in the HE where A copies are overexpressed in *Yudal* and C copies overexpressed in *Darmor-bzh*. **B) Fixed HEs; Darmor/Yudal**: The fold-change (log2) for *Darmor-bzh* / *Yudal* expression, for genes on the C genome. Clusters of genes differentially expressed between *Darmor-bzh* and *Yudal* appear as distinct peaks. **C) Newly formed HE; Darmor ACCC/AACC and CCCC/AACC**: The fold-change (log2) for ACCC/AACC expression (grey dots) and CCCC/AACC expression (black dots) for genes on chromosome A01. A large terminal cluster of differentially expressed genes is evident. **D) Total$_{(A+C)}$ expression following newly formed (left) and fixed (right) HEs:** The ratio of post-HE Total$_{(A+C)}$ expression (i.e. A+C in CCCC or AAAA) over pre-HE Total$_{(A+C)}$ expression (i.e. A+C in AACC), for genes showing increasing contribution of the duplicated homoeologue to pre-HE Total$_{(A+C)}$ expression. **For the new HE in *Darmor-bzh***: The best fit regression line (in blue) for the data (blue dots) is almost identical to the black line (post-HE Total$_{(A+C)}$ expression = 2 × pre-HE expression of the donor homoeologue) but different from the red line (constant Total$_{(A+C)}$ expression). **For the fixed HE in *Yudal*:** The best fit regression line (blue) is in-between the black and red lines; i.e. post-HE Total$_{(A+C)}$ expression in *Yudal* approaches pre-HE Total$_{(A+C)}$ inferred from *Darmor-bzh* Total$_{(A+C)}$. **E) Post-HE transcriptional compensation:** The ratio of post-HE Total$_{(A+C)}$ expression over pre-HE Total$_{(A+C)}$ for genes with a large homoeologue bias (Low or High; A:C log2 ratio < -2 or > 2 respectively) in the newly formed (New) and fixed (Old) HEs. **For the new HE** no compensation is seen and expression is additive (Total$_{(A+C)}$ Observed:Expected log2.ratio = 0). **For fixed HEs**, compensation is observed: duplication of the lower expressed copy results in higher than expected expression (Total$_{(A+C)}$ Observed:Expected log2.ratio > 0; Total$_{(A+C)}$ Observed:Expected HE > NonHE; * $p < 0.05$, *** $p < 0.001$).

Table 1. Expression biases for genes within fixed *Yudal* HEs

	Homoeologue bias prior to HE (*Darmor-bzh* - A/[A+C])			
Yudal$_{(A+C)}$ vs *Darmor-bzh*$_{(A+C)}$	A < C	A ~ C	A > C	p-value*
$Y_{(A+C)} < D_{(A+C)}$	38	19	35	0.73
$Y_{(A+C)} \sim D_{(A+C)}$	25	26	18	0.29
$Y_{(A+C)} > D_{(A+C)}$	17	24	61	3.59E-07
Total	80	69	114	0.015

*Binomial test: (A < C) = (A > C)

Materials and methods

Plant material

We used inbred genotypes (i.e. maintained by selfing) from *Brassica napus* cv *Darmor-bzh*, a French dwarf winter cultivar and *B.napus* cv *Yudal*, a spring korean line. These same two varieties were used by Chalhoub et al., (2014) for reference genome sequencing (*Darmor-bzh*) and for resequencing (*Yudal*), respectively.

Analysis of homoeologous exchanges.

We used polymerase chain reaction (PCR) with consensus primers pairs between homoeologous copies (Table S1) for genes falling within the borders of inferred HEs. The presence of homoeologous single nucleotide polymorphisms (homoeo-SNPs) or small insertion-deletion was revealed by Sanger sequencing (Fig. S1). We also used the 15K Brassica array from TraitGenetics to infer additional markers within HEs.

Recombination Frequency and Gene Density.

Recombination frequency of the average chromosome arm was determined from the *Darmor-bzh* x *Yudal* recombination data reported in Delourme et al. (Delourme et al., 2013). Gene density was the proportion of sequence annotated "gene" in the *Darmor-bzh* genome annotation file (Chalhoub et al., 2014). Centromere positions used were those reported by Mason et al. (Mason et al., 2015).

Male meiocyte collection, RNA extraction and Sequencing.

Pollen mother cells were isolated as described in Chen et al. (Chen et al., 2010) and used for RNA extraction. cDNA libraries were prepared using TruSeq technology and sequenced on an Illumina Hiseq2000 in paired-end. After trimming, 30-40 million 100-bp long reads were used for every sample (Table S9).

Experimental design.

For both *Darmor-bzh* and *Yudal*, three biological replicates (i.e. three plants) and two technical replicates per plant were used to prepare a total of 12 cDNA libraries. The two libraries from each plant were divided in two parts and sequenced on a different lane. As a consequence, each plant was represented on each lane (4 lanes in total).

Read mapping.

Allelic SNP-tolerant mapping of reads onto the *B. napus* (var. *Darmor-bzh*) reference genome was performed using GSNAP (v 2013-10-28). Only reads that mapped uniquely to the A or to the C genomes were considered for further analyses. Count matrixes were obtained using htseq-count, considering each gene a transcription unit.

Statistical Analyses.

Statistical analyses were carried out using the subset of genes that were transcribed in meiocytes above a threshold of 10 reads across the 12 samples.

Confirmation of HEs' divergent expression profiles

To test whether the fixed HEs in *Darmor-bzh* and *Yudal* showed a specific transcriptome profile, we assessed whether the duplicated and lost regions associated with an HE containing *n* genes was enriched in up-regulated or down-regulated genes respectively, compared to a) all contiguous regions of *n* genes or b) sets of *n* randomly chosen genes (Table S4).

de novo detection of HEs

For *de novo* detection of HEs we undertook segmentation analyses as described in Rigail et al., (2012) to identify genomic segments characterized by a constant expression profile (i.e. up- or down-regulated in *Darmor-bzh* compared to *Yudal*); these segments were then compared to the position of fixed HEs (Table S5-6). The same analysis was used for *de novo* identification of HEs in the three *Darmor-bzh* (or *Yudal*) biological replicates (hereafter D1, D2 and D3) to identify regions where the expression ratio of D1 over D2, D1 over D3 and D2 over D3 is not equal to 1. In addition to using the coordinates of the genes along the *B. napus* pseudo-molecules (Chalhoub et al., 2014) we also ran the analyses using the coordinates of their orthologues in *B. rapa* (v1.5) in order to integrate genes that are within well assembled regions in the *B. rapa* assembly but allocated to non-anchored scaffolds (chrX_random) in the *B. napus* assembly.

Assessing proportional effect of HEs on gene expression

To test whether gene copy number had a proportional effect on gene expression we used Deseq2 (Love et al., 2014). We considered several normalization approaches (Figure SX): a) the **S. Fact Int** approach corresponded to a standard deseq2 analysis; we computed the size factors using the genes contained within HEs. With this first approach, we expected few differentially

expressed genes as the internal calculation of the size factors should compensate for any DNA copy number effect on gene expression. b) In the **S. Fact Ext** approach, we computed the size factors using genes outside HEs. With this second approach, many differentially expressed genes are expected (e.g. expression of A genes in D2 [2A] is expected to be higher than in D3 [1A]). c) In the **S. Fact Ext * copy** approach we also computed the size factors using genes outside of the new HE but then multiplied these values by the copy number (e.g. 2 for sample D2 and 1 for sample D3). With this third analysis, very few differences are expected if expression is proportional to copy number for most genes.

Assessment of gene content and transcriptional output of HEs

To identify the three categories of genes in the AACC genotype (A > C, A = C, A < C) we used Deseq2 (Love et al., 2014). We also used Deseq2 to identify whether $Total_{(A+C)}$ expression in the AAAA (or CCCC) genotype was greater than, equal to, or less than, $Total_{(A+C)}$ expression in the AACC genotype. For each $Total_{(A+C)}$ expression category we determined whether the genes were split equally across the A/C ratio categories using a binomial test (A < C = A > C). We also tested whether the proportion of genes within each of the A/C ratio categories for HEs differed from that of genes outside HEs using a Fisher exact test.

More details of statistical analyses are given in SM - Statistics.

Acknowledgments.
We would like to thank M. Grelon, S. Nicolas, C. Mézard and K. Alix for critical reading and discussion of the manuscript. This work has been funded through the ANR project ANR-2010-BLANC-628 SVSE7- DUPLIC and a grant from the Biology and Plant Breeding division of INRA (AAP-GAP-2012-VARCO). The IJPB benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS). A.B. is funded by a "Young Scientist Contracts" (CJS) from INRA. A.L. is currently funded by a Marie Curie postdoctoral fellowship (PIOF-GA-2013-628128, PolyMeio).

Author contributions
A.L., D.C., E.D. and E.J. designed the research, A.L., A.B., D.C., C.C., J.T., S.B., performed the research, A.L., A.B., D.C., G.R. and E.J. analyzed the data, A.L., A.B., G.R. and E.J. wrote the paper.

Competing financial interests.
The authors declare no competing financial interests.

References

Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U. S. A. 100:4649–4654.

Albalat R, Cañestro C. 2016. Evolution by gene loss. Nat. Rev. Genet. 17:379–391.

Bai Z, Chen J, Liao Y, Wang M, Liu R, Ge S, Wing RA, Chen M, Girirajan S, Campbell C, et al. 2016. The impact and origin of copy number variations in the Oryza species. BMC Genomics 17:261.

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, et al. 2016. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat. Genet. 48:438–446.

Butruille D V, Guries RP, Osborn TC, Alpert KB, Tanksley SD, Azanza F, Young TE, Kim D, Tanksley SD, Juvik JA, et al. 1999. Linkage analysis of molecular markers and quantitative trait loci in populations of inbred backcross lines of Brassica napus L. Genetics 153:949–964.

Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. 17:224–238.

Chalhoub B, Denoeud F, Liu S, Parkin I a. P, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. 2014. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345:950–953.

Chen C, Farmer AD, Langley RJ, Mudge J, Crow J a, May GD, Huntley J, Smith AG, Retzel EF. 2010. Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. 10:280.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev E V, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proc. Natl. Acad. Sci. U. S. A. 109:1176–1181.

Delourme R, Falentin C, Fomeju BF, Boillot M, Lassalle G, André I, Duarte J, Gauthier V, Lucante N, Marty A, et al. 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in Brassica napus L. BMC Genomics 14:120.

Feuillet C, Penger A, Gellner K, Mast A, Keller B. 2001. Molecular Evolution of Receptor-Like Kinase Genes in Hexaploid Wheat. Independent Evolution of Orthologs after Polyploidization and Mechanisms of Local Rearrangements at Paralogous Loci. PLANT Physiol. 125:1304–1313.

Gaeta RT, Chris Pires J. 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol. 186:18–28.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic Changes in Resynthesized Brassica napus and Their Effect on Gene Expression and Phenotype. PLANT CELL ONLINE 19:3403–3417.

Gaeta RT, Yoo S-Y, Pires JC, Doerge RW, Chen ZJ, Osborn TC. 2009. Analysis of Gene Expression in Resynthesized Brassica napus Allopolyploids Using Arabidopsis 70mer Oligo Microarrays.Hazen SP, editor. PLoS One 4:e4760.

Gong L, Olson M, Wendel JF. 2014. Cytonuclear evolution of rubisco in four allopolyploid lineages. Mol. Biol. Evol. 31:2624–2636.

Grandont L, Cuñado N, Coriton O, Huteau V, Eber F, Chèvre AM, Grelon M, Chelysheva L, Jenczewski E. 2014. Homoeologous chromosome sorting and progression of meiotic recombination in *Brassica napus*: ploidy does matter! Plant Cell.

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G. 2006. Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752.

Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SAAC, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD, et al. 2008. Distribution and functional impact of DNA copy number

variation in the rat. Nat. Genet. 40:538–545.
Henrichsen CN, Chaignat E, Reymond A. 2009. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18:R1-8.
Henry IM, Dilkes BP, Tyagi A, Gao J, Christensen B, Comai L. 2014. The *BOY NAMED SUE* quantitative trait locus confers increased meiotic stability to an adapted natural allopolyploid of *Arabidopsis*. Plant Cell 26:181–194.
Hose J, Yong CM, Sardi M, Wang Z, Newton MA, Gasch AP. 2015. Dosage compensation can buffer copy-number variation in wild yeast. Elife 4.
Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ. 2008. A and C genome distinction and chromosome identification in brassica napus by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics 180:1849–1857.
International Wheat Genome Sequencing Consortium (IWGSC) TIWGSC, Lobell DB, Schlenker W, Costa-Roberts J, Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S, Foley JA, et al. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345:1251788.
Lashermes P, Combes MC, Hueber Y, Severac D, Dereeper A. 2014. Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. Plant J. 78:674–685.
Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, et al. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. 33:524–530.
Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. 2016. Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. Plant Cell 28:326–344.
Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR. 2008. Rapid Chromosome Evolution in Recently Formed Polyploids in Tragopogon (Asteraceae). PLoS One 3:e3353.
Lloyd AH. 2014. Meiotic gene evolution: can you teach a new dog new tricks? Mol Biol Evol.
Lombard V, Delourme R. 2001. A consensus linkage map for rapeseed (Brassica napus L.): construction and integration of three individual maps from DH populations. TAG Theor. Appl. Genet. 103:491–507.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.
Mason AS, Rousseau-gueutin M, Morice J, Bayer PE. 2015. Centromere Locations in Brassica A and C Genomes Revealed Through Half- Tetrad Analysis. 202:1–31.
Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, et al. 2007. Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of Brassica napus haploids. Genetics 175:487–503.
Nicolas SD, Monod H, Eber F, Chèvre A-M, Jenczewski E. 2012. Non-random distribution of extensive chromosome rearrangements in Brassica napus depends on genome organization. Plant J. 70:691–703.
Osborn TC, Butrulle D V, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiate DJ. 2003. Detection and Effects of a Homeologous Reciprocal Transposition in Brassica napus. Genetics 165:1569–1577.
Pala I, Coelho MM, Schartl M, Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S, Otto SP, Whitton J, et al. 2008. Dosage compensation by gene-copy silencing in a triploid hybrid fish. Curr. Biol. 18:1344–1348.
Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Cáceres AM, Iafrate AJ, Tyler-Smith C, Scherer SW, Eichler EE, et al. 2006. Hotspots for copy number variation in chimpanzees and humans. Proc. Natl. Acad. Sci. U. S. A. 103:8006–8011.
Pezer Ž, Harr B, Teschke M, Babiker H, Tautz D. 2015. Divergence patterns of genic copy number variation in natural populations of the house mouse (Mus musculus domesticus)

reveal three conserved genes with major population-specific expansions. Genome Res. 25:1114–1124.
Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, doucet I, Perret D, Villeger M-J, Vincourt P, Blanchard P. 2005. Construction of an oilseed rape (Brassica napus L.) genetic map with SSR markers. Theor. Appl. Genet. 111:1514–1523.
Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada A, Lukens LN, Osborn TC. 2004. Biological relevance of polyploidy : ecology to genomics Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). :675–688.
Saxena RK, Edwards D, Varshney RK. 2014. Structural variations in plant genomes. Brief. Funct. Genomics 13:296–307.
Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2015. Cytonuclear Coordination Is Not Immediate upon Allopolyploid Formation in Tragopogon miscellus (Asteraceae) Allopolyploids. PLoS One 10:e0144339.
Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ. 1995. Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (Brassica napus). Genome 38:1112–1121.
Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre AM. 2011. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized Brassica napus. New Phytol. 191:884–894.
Udall JA, Quijada PA, Osborn TC. 2005. Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of Brassica napus L. Genetics 169:967–979.
Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172:507–517.
Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat. Rev. Genet. 14:125–138.
Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol. 8:e1000409.
Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc. Natl. Acad. Sci. U. S. A. 108:7908–7913.
Yoo M-J, Liu X, Pires JC, Soltis PS, Soltis DE. 2014. Nonadditive gene expression in polyploids. Annu. Rev. Genet. 48:485–517.
Zhang Y, Oliver B. 2007. Dosage compensation goes global. Curr. Opin. Genet. Dev. 17:113–120.
Zhang Z, Mao L, Chen H, Bu F, Li G, Sun J, Li S, Sun H, Jiao C, Blakely R, et al. 2015. Genome-Wide Mapping of Structural Variations Reveals a Copy Number Variant That Determines Reproductive Morphology in Cucumber. Plant Cell 27:1595–1604.
Zhou R, Moshgabadi N, Adams KL. 2011. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc. Natl. Acad. Sci. 108:16122–16127.

Annexes

Supplementary Data

Homoeologous Exchanges in the reference genome hider short read mapping

We analyzed the impact of non-shared HEs between *Darmor-bzh* (the reference genotype used for genome assembly) and *Yudal* (serving as an example for non–reference genotypes) on RNAseq or other short-reads mapping. We anticipated that reads coming from a region that is only present in *Yudal* (as a result of a HE in *Darmor-bzh*) cannot be mapped when mapping parameters are set to discriminate homoeologous copies. To test this possibility, we relaxed the mapping stringency parameters for *Yudal* and identified windows of the pseudomolecules where the number of mapped reads from *Yudal* consistently increased. This approach was expected to identify HEs that are unique to *Darmor-bzh* and indeed validated most of the unique *Darmor-bzh* HEs that we had previously confirmed by PCR (Figure 1A). This approach did not identify the smallest events in *Darmor-bzh* as well as event 1DAn2+ / 1DCn2- that is not unique to *Darmor-bzh* but overlap with 1YAn2+ / 1YCn2- in *Yudal* (Figure 1A).

Nineteen other regions of increased read mapping were identified in addition to the 9 previously confirmed unique *Darmor-bzh* HEs. These regions do not correspond to HEs in *Darmor-bzh* as their homoeologous counterparts were always present in the *B. napus* assembly. Seven of these regions overlapped with HEs in *Yudal*, in particular with the associated duplicated regions (5 out of 7).

Supplementary Figures

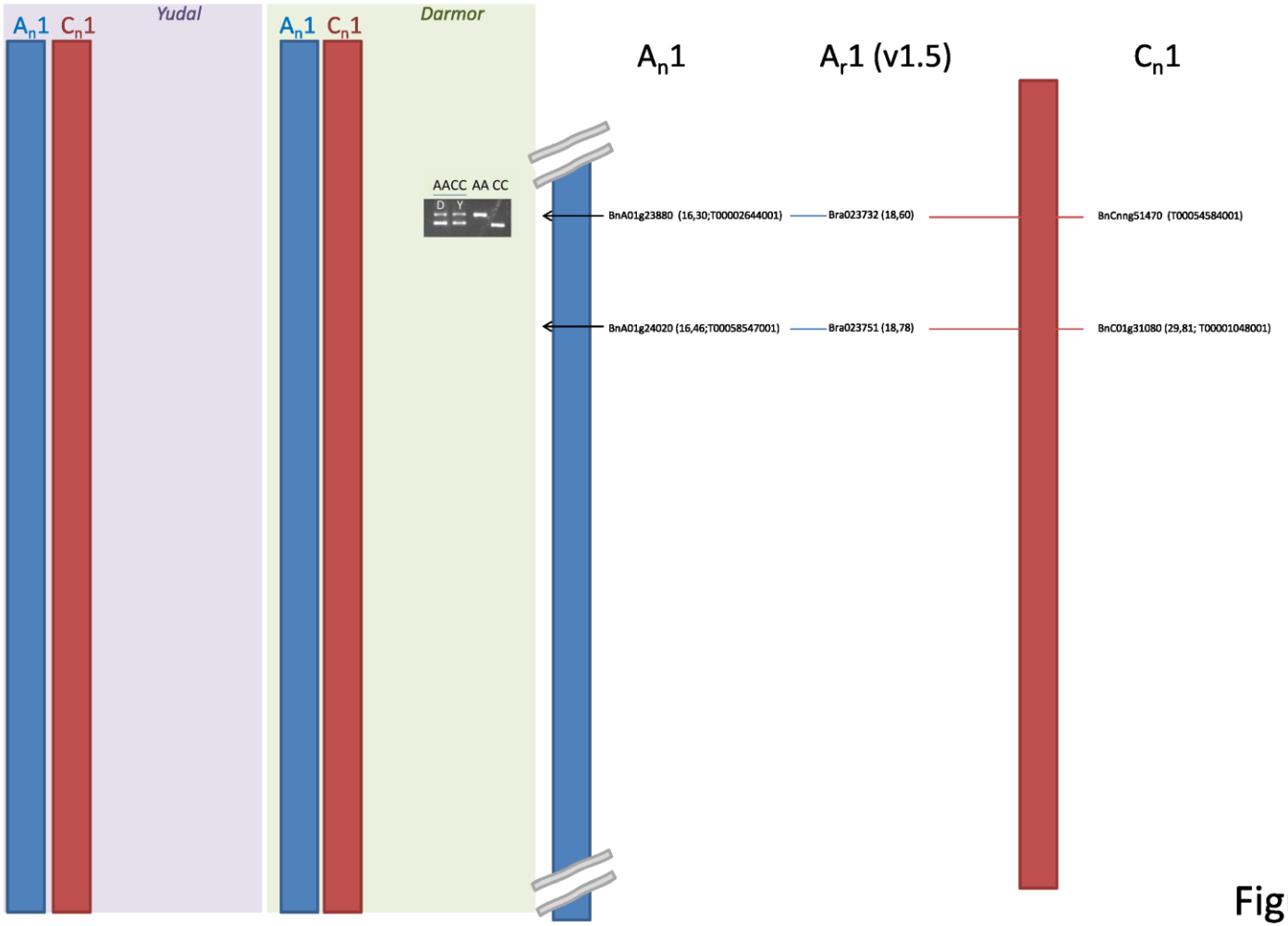


Figure S1A

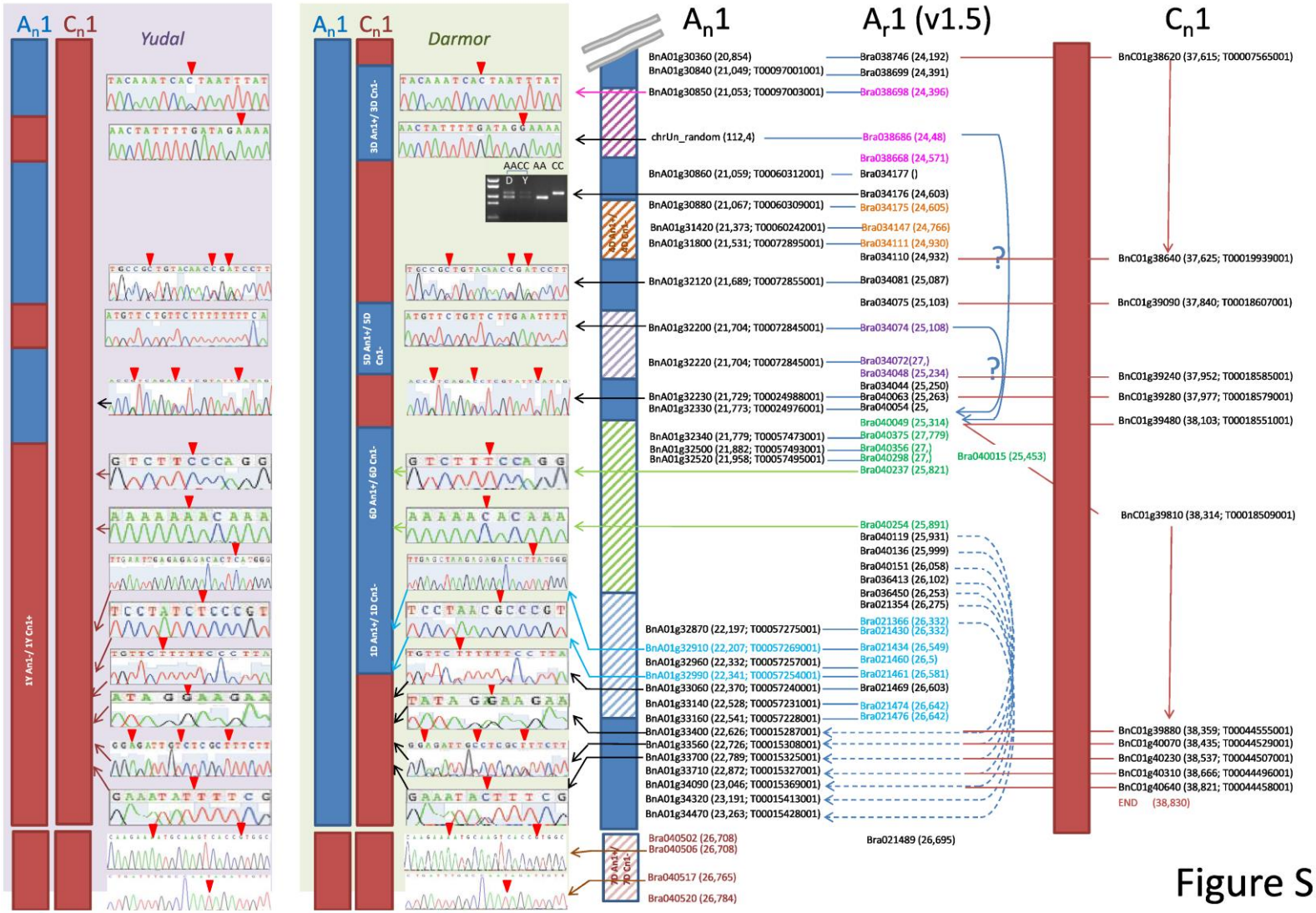

An1 Cn1
Yudal
An1 Cn1
Darmor
AACC AA CC
D Y
An1
Ar1 (v1.5)
Cn1
BnA01g30360 (20,854)
BnA01g30840 (21,049; T00097001001)
BnA01g30850 (21,053; T00097003001)
chrUn_random (112,4)
BnA01g30860 (21,059; T00060312001)
BnA01g30880 (21,067; T00060309001)
BnA01g31420 (21,373; T00060242001)
BnA01g31800 (21,531; T00072895001)
BnA01g32120 (21,689; T00072855001)
BnA01g32200 (21,704; T00072845001)
BnA01g32220 (21,704; T00072845001)
BnA01g32230 (21,729; T00024988001)
BnA01g32330 (21,773; T00024976001)
BnA01g32340 (21,779; T00057473001)
BnA01g32500 (21,882; T00057493001)
BnA01g32520 (21,958; T00057495001)
BnA01g32870 (22,197; T00057275001)
BnA01g32910 (22,207; T00057269001)
BnA01g32960 (22,332; T00057257001)
BnA01g32990 (22,341; T00057254001)
BnA01g33060 (22,370; T00057240001)
BnA01g33140 (22,528; T00057231001)
BnA01g33160 (22,541; T00057228001)
BnA01g33400 (22,626; T00015287001)
BnA01g33560 (22,726; T00015308001)
BnA01g33700 (22,789; T00015325001)
BnA01g33710 (22,872; T00015327001)
BnA01g34090 (23,046; T00015369001)
BnA01g34320 (23,191; T00015413001)
BnA01g34470 (23,263; T00015428001)
Bra038746 (24,192)
Bra038699 (24,391)
Bra038698 (24,396)
Bra038686 (24,48)
Bra038668 (24,571)
Bra034177 ()
Bra034176 (24,603)
Bra034175 (24,605)
Bra034147 (24,766)
Bra034111 (24,930)
Bra034110 (24,932)
Bra034081 (25,087)
Bra034075 (25,103)
Bra034074 (25,108)
Bra034072(27,)
Bra034048 (25,234)
Bra034044 (25,250)
Bra040063 (25,263)
Bra040054 (25,
Bra040049 (25,314)
Bra040375 (27,779)
Bra040356 (27,)
Bra040298 (27,)
Bra040237 (25,821)
Bra040254 (25,891)
Bra040119 (25,931)
Bra040136 (25,999)
Bra040151 (26,058)
Bra036413 (26,102)
Bra036450 (26,253)
Bra021354 (26,275)
Bra021366 (26,332)
Bra021430 (26,332)
Bra021434 (26,549)
Bra021460 (26,5)
Bra021461 (26,581)
Bra021469 (26,603)
Bra021474 (26,642)
Bra021476 (26,642)
Bra021489 (26,695)
Bra040502 (26,708)
Bra040506 (26,708)
Bra040517 (26,765)
Bra040520 (26,784)
Bra040015 (25,453)
BnC01g38620 (37,615; T00007565001)
BnC01g38640 (37,625; T00019939001)
BnC01g39090 (37,840; T00018607001)
BnC01g39240 (37,952; T00018585001)
BnC01g39280 (37,977; T00018579001)
BnC01g39480 (38,103; T00018551001)
BnC01g39810 (38,314; T00018509001)
BnC01g39880 (38,359; T00044555001)
BnC01g40070 (38,435; T00044529001)
BnC01g40230 (38,537; T00044507001)
BnC01g40310 (38,666; T00044496001)
BnC01g40640 (38,821; T00044458001)
END (38,830)
?
?

Figure S1B

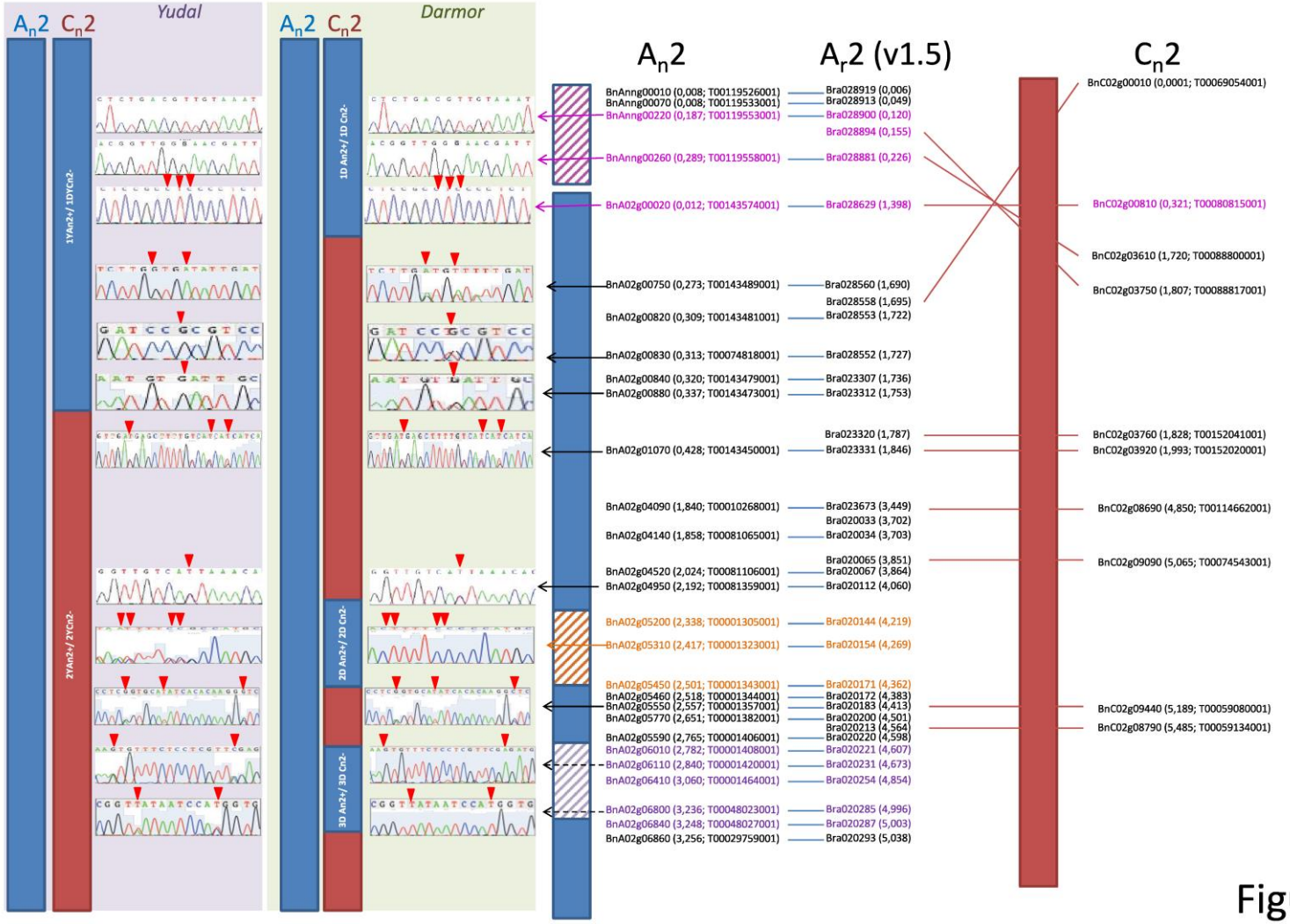

Yudal
Darmor
An2
Cn2
1YAn2+/ 1DYCn2-
2YAn2+/ 2YCn2-
1D An2+/ 1D Cn2-
2D An2+/ 2D Cn2-
3D An2+/ 3D Cn2-
An2
Ar2 (v1.5)
Cn2
BnAnng00010 (0,008; T00119526001)
BnAnng00070 (0,008; T00119533001)
BnAnng00220 (0,187; T00119553001)
BnAnng00260 (0,289; T00119558001)
BnA02g00020 (0,012; T00143574001)
BnA02g00750 (0,273; T00143489001)
BnA02g00820 (0,309; T00143481001)
BnA02g00830 (0,313; T00074818001)
BnA02g00840 (0,320; T00143479001)
BnA02g00880 (0,337; T00143473001)
BnA02g01070 (0,428; T00143450001)
BnA02g04090 (1,840; T00010268001)
BnA02g04140 (1,858; T00081065001)
BnA02g04520 (2,024; T00081106001)
BnA02g04950 (2,192; T00081359001)
BnA02g05200 (2,338; T00001305001)
BnA02g05310 (2,417; T00001323001)
BnA02g05450 (2,501; T00001343001)
BnA02g05460 (2,518; T00001344001)
BnA02g05550 (2,557; T00001357001)
BnA02g05770 (2,651; T00001382001)
BnA02g05590 (2,765; T00001406001)
BnA02g06010 (2,782; T00001408001)
BnA02g06110 (2,840; T00001420001)
BnA02g06410 (3,060; T00001464001)
BnA02g06800 (3,236; T00048023001)
BnA02g06840 (3,248; T00048027001)
BnA02g06860 (3,256; T00029759001)
Bra028919 (0,006)
Bra028913 (0,049)
Bra028900 (0,120)
Bra028894 (0,155)
Bra028881 (0,226)
Bra028629 (1,398)
Bra028560 (1,690)
Bra028558 (1,695)
Bra028553 (1,722)
Bra028552 (1,727)
Bra023307 (1,736)
Bra023312 (1,753)
Bra023320 (1,787)
Bra023331 (1,846)
Bra023673 (3,449)
Bra020033 (3,702)
Bra020034 (3,703)
Bra020065 (3,851)
Bra020067 (3,864)
Bra020112 (4,060)
Bra020144 (4,219)
Bra020154 (4,269)
Bra020171 (4,362)
Bra020172 (4,383)
Bra020183 (4,413)
Bra020200 (4,501)
Bra020213 (4,564)
Bra020220 (4,598)
Bra020221 (4,607)
Bra020231 (4,673)
Bra020254 (4,854)
Bra020285 (4,996)
Bra020287 (5,003)
Bra020293 (5,038)
BnC02g00010 (0,0001; T00069054001)
BnC02g00810 (0,321; T00080815001)
BnC02g03610 (1,720; T00088800001)
BnC02g03750 (1,807; T00088817001)
BnC02g03760 (1,828; T00152041001)
BnC02g03920 (1,993; T00152020001)
BnC02g08690 (4,850; T00114662001)
BnC02g09090 (5,065; T00074543001)
BnC02g09440 (5,189; T00059080001)
BnC02g08790 (5,485; T00059134001)

Figure S1C

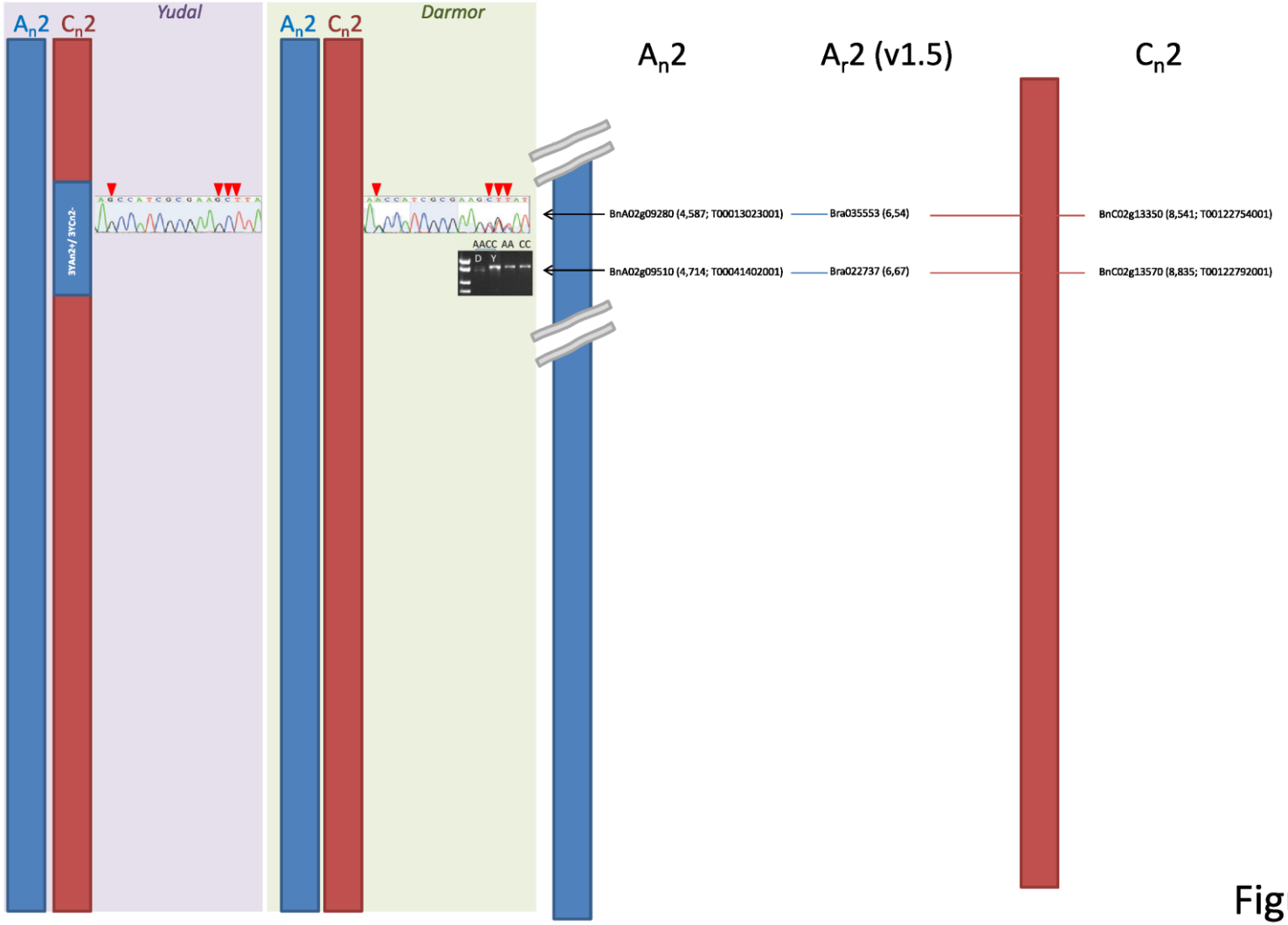

Yudal
Darmor
A_n2 C_n2
A_n2
A_r2 (v1.5)
C_n2
3YAn2+/ 3YCn2-
AACC AA CC
D Y
BnA02g09280 (4,587; T00013023001)
Bra035553 (6,54)
BnC02g13350 (8,541; T00122754001)
BnA02g09510 (4,714; T00041402001)
Bra022737 (6,67)
BnC02g13570 (8,835; T00122792001)

Figure S1D

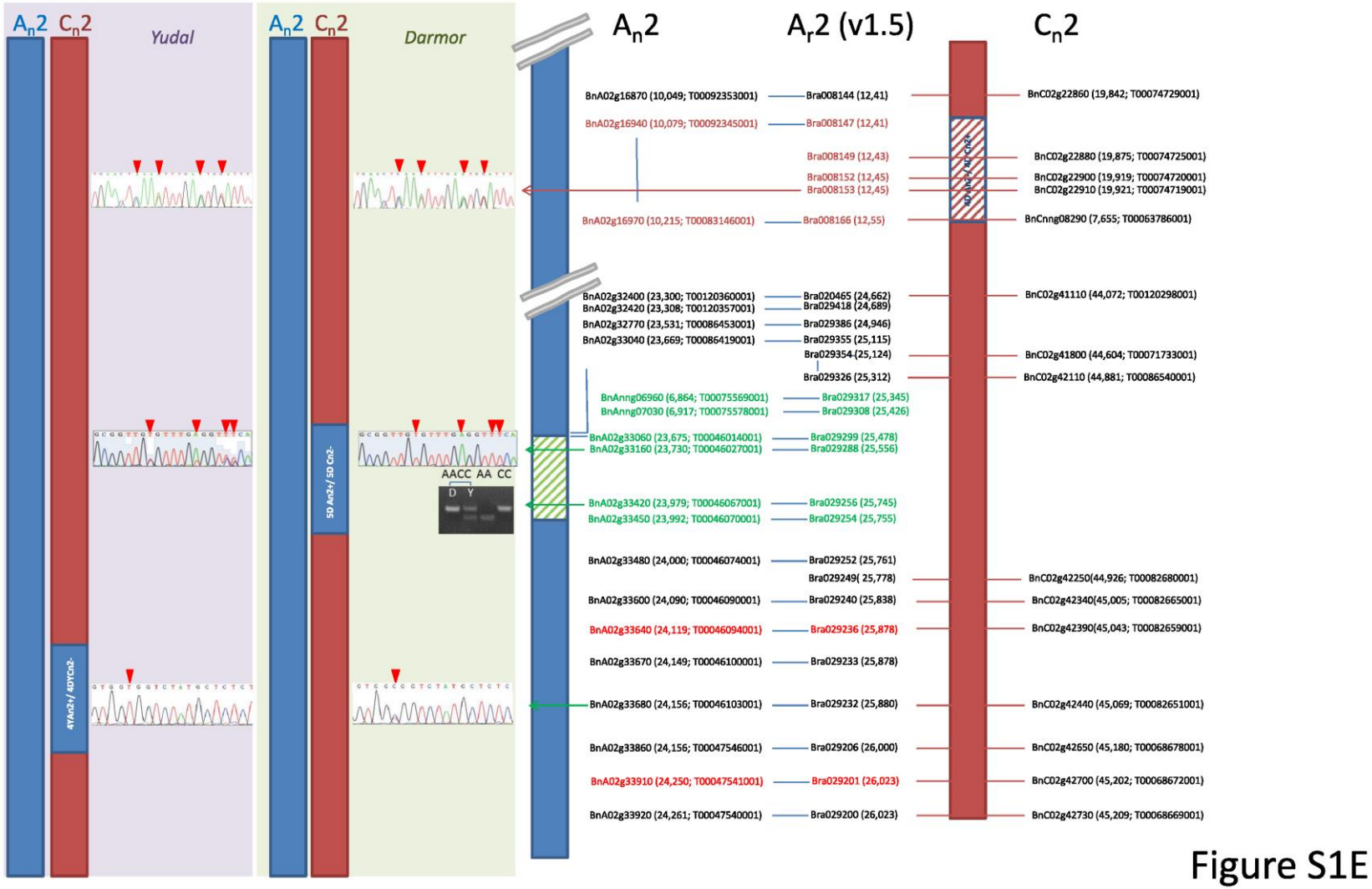

An2 Cn2
Yudal
An2 Cn2
Darmor
An2
Ar2 (v1.5)
Cn2
BnA02g16870 (10,049; T00092353001)
Bra008144 (12,41)
BnC02g22860 (19,842; T00074729001)
BnA02g16940 (10,079; T00092345001)
Bra008147 (12,41)
Bra008149 (12,43)
BnC02g22880 (19,875; T00074725001)
Bra008152 (12,45)
BnC02g22900 (19,919; T00074720001)
Bra008153 (12,45)
BnC02g22910 (19,921; T00074719001)
BnA02g16970 (10,215; T00083146001)
Bra008166 (12,55)
BnCnng08290 (7,655; T00063786001)
BnA02g32400 (23,300; T00120360001)
Bra020465 (24,662)
BnC02g41110 (44,072; T00120298001)
BnA02g32420 (23,308; T00120357001)
Bra029418 (24,689)
BnA02g32770 (23,531; T00086453001)
Bra029386 (24,946)
BnA02g33040 (23,669; T00086419001)
Bra029355 (25,115)
Bra029354 (25,124)
BnC02g41800 (44,604; T00071733001)
Bra029326 (25,312)
BnC02g42110 (44,881; T00086540001)
BnAnng06960 (6,864; T00075569001)
Bra029317 (25,345)
BnAnng07030 (6,917; T00075578001)
Bra029308 (25,426)
BnA02g33060 (23,675; T00046014001)
Bra029299 (25,478)
BnA02g33160 (23,730; T00046027001)
Bra029288 (25,556)
AACC AA CC
D Y
BnA02g33420 (23,979; T00046067001)
Bra029256 (25,745)
BnA02g33450 (23,992; T00046070001)
Bra029254 (25,755)
BnA02g33480 (24,000; T00046074001)
Bra029252 (25,761)
Bra029249(25,778)
BnC02g42250(44,926; T00082680001)
BnA02g33600 (24,090; T00046090001)
Bra029240 (25,838)
BnC02g42340(45,005; T00082665001)
BnA02g33640 (24,119; T00046094001)
Bra029236 (25,878)
BnC02g42390(45,043; T00082659001)
BnA02g33670 (24,149; T00046100001)
Bra029233 (25,878)
BnA02g33680 (24,156; T00046103001)
Bra029232 (25,880)
BnC02g42440 (45,069; T00082651001)
BnA02g33860 (24,156; T00047546001)
Bra029206 (26,000)
BnC02g42650 (45,180; T00068678001)
BnA02g33910 (24,250; T00047541001)
Bra029201 (26,023)
BnC02g42700 (45,202; T00068672001)
BnA02g33920 (24,261; T00047540001)
Bra029200 (26,023)
BnC02g42730 (45,209; T00068669001)

Figure S1E

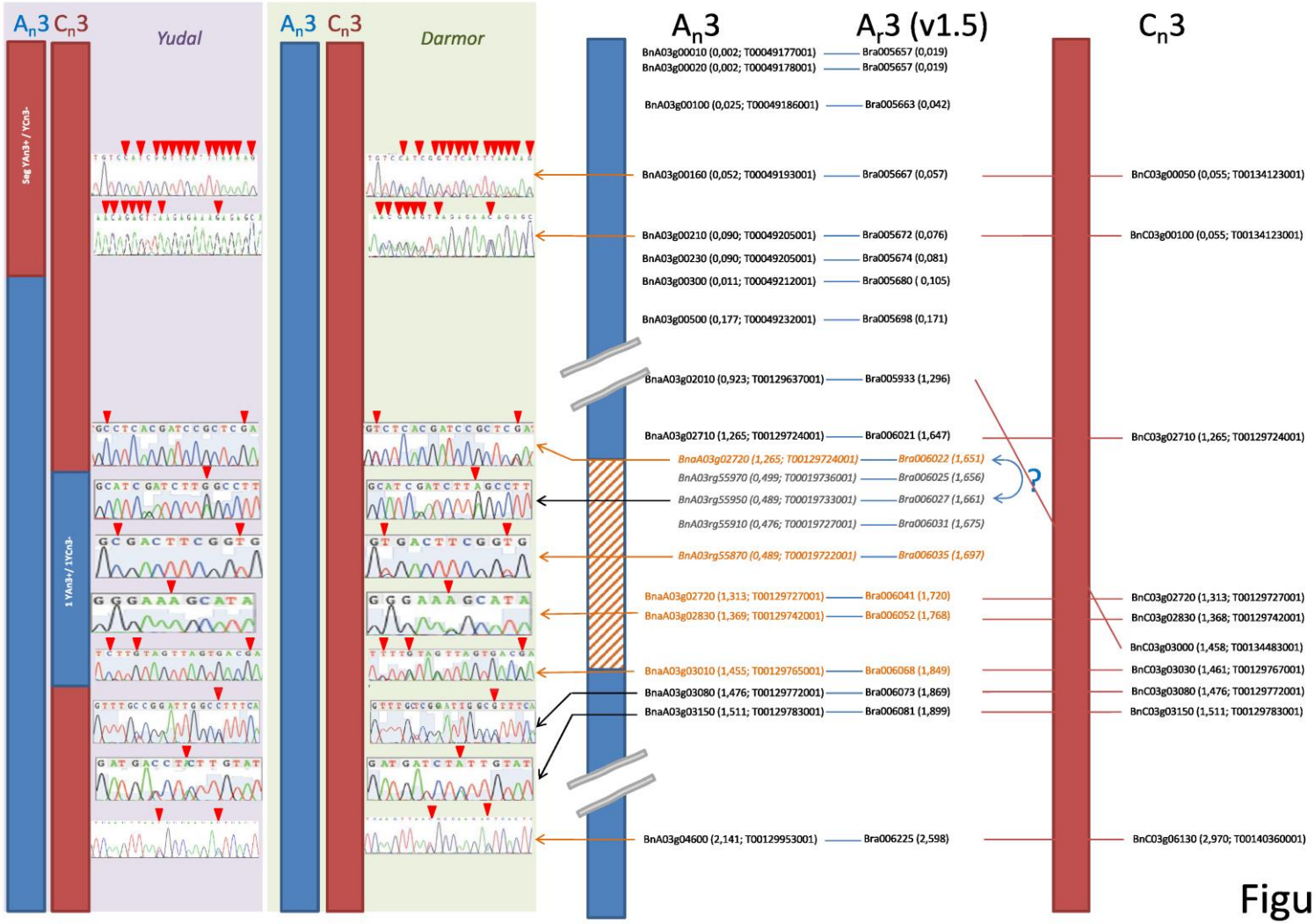

An3 Cn3
Yudal
An3 Cn3
Darmor
An3
Ar3 (v1.5)
Cn3
BnA03g00010 (0,002; T00049177001) Bra005657 (0,019)
BnA03g00020 (0,002; T00049178001) Bra005657 (0,019)
BnA03g00100 (0,025; T00049186001) Bra005663 (0,042)
BnA03g00160 (0,052; T00049193001) Bra005667 (0,057)
BnC03g00050 (0,055; T00134123001)
BnA03g00210 (0,090; T00049205001) Bra005672 (0,076)
BnC03g00100 (0,055; T00134123001)
BnA03g00230 (0,090; T00049205001) Bra005674 (0,081)
BnA03g00300 (0,011; T00049212001) Bra005680 (0,105)
BnA03g00500 (0,177; T00049232001) Bra005698 (0,171)
BnaA03g02010 (0,923; T00129637001) Bra005933 (1,296)
BnaA03g02710 (1,265; T00129724001) Bra006021 (1,647)
BnC03g02710 (1,265; T00129724001)
BnaA03g02720 (1,265; T00129724001) Bra006022 (1,651)
BnA03rg55970 (0,499; T00019736001) Bra006025 (1,656)
BnA03rg55950 (0,489; T00019733001) Bra006027 (1,661)
BnA03rg55910 (0,476; T00019727001) Bra006031 (1,675)
BnA03rg55870 (0,489; T00019722001) Bra006035 (1,697)
?
BnaA03g02720 (1,313; T00129727001) Bra006041 (1,720)
BnC03g02720 (1,313; T00129727001)
BnaA03g02830 (1,369; T00129742001) Bra006052 (1,768)
BnC03g02830 (1,368; T00129742001)
BnC03g03000 (1,458; T00134483001)
BnaA03g03010 (1,455; T00129765001) Bra006068 (1,849)
BnC03g03030 (1,461; T00129767001)
BnaA03g03080 (1,476; T00129772001) Bra006073 (1,869)
BnC03g03080 (1,476; T00129772001)
BnaA03g03150 (1,511; T00129783001) Bra006081 (1,899)
BnC03g03150 (1,511; T00129783001)
BnA03g04600 (2,141; T00129953001) Bra006225 (2,598)
BnC03g06130 (2,970; T00140360001)

Figure S1F

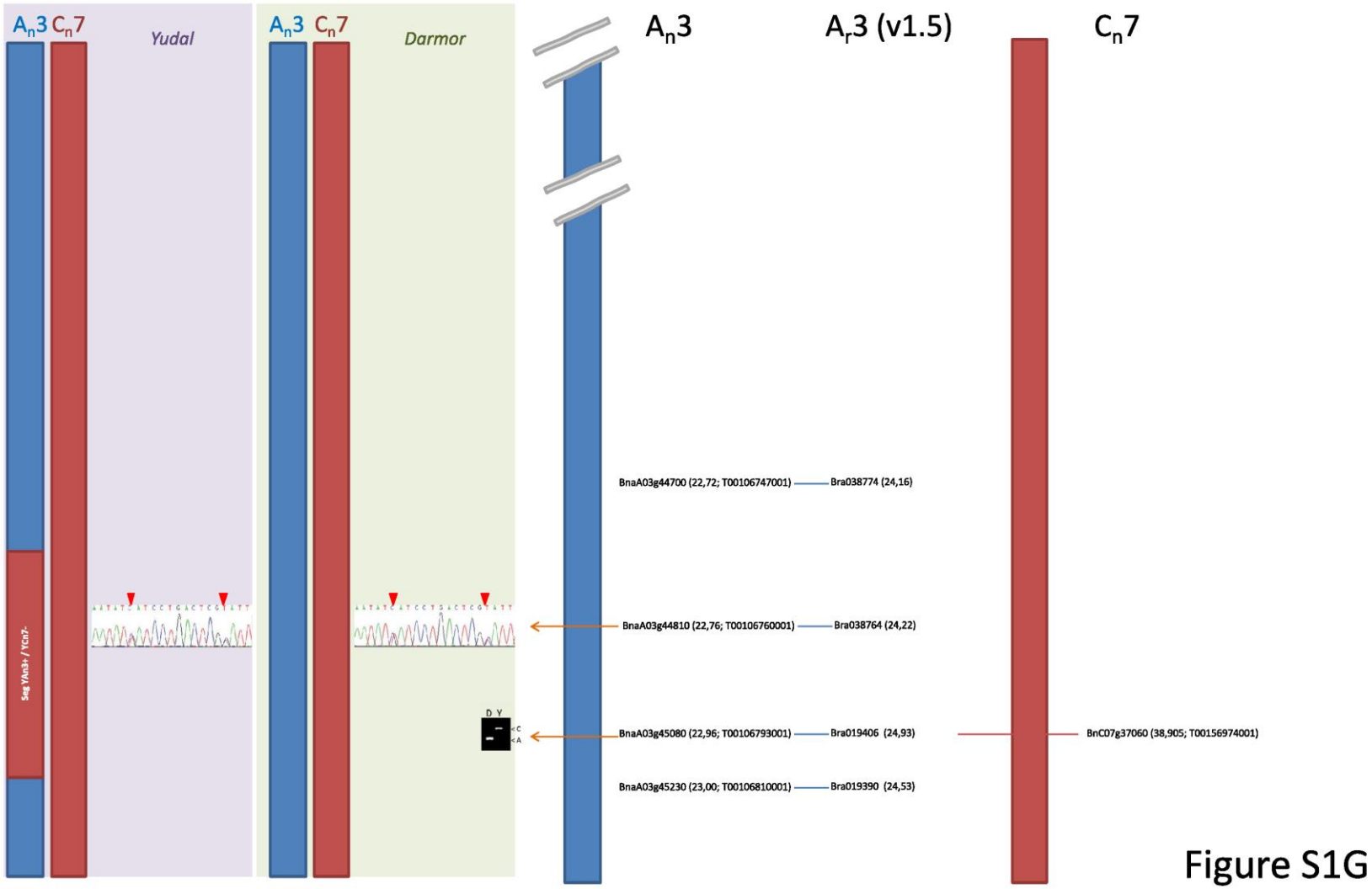

An3 Cn7
Yudal
An3 Cn7
Darmor
An3
Ar3 (v1.5)
Cn7
Seg YAn3+ / YCn7-
BnaA03g44700 (22,72; T00106747001)
Bra038774 (24,16)
BnaA03g44810 (22,76; T00106760001)
Bra038764 (24,22)
D Y
BnaA03g45080 (22,96; T00106793001)
Bra019406 (24,93)
BnaA03g45230 (23,00; T00106810001)
Bra019390 (24,53)
BnC07g37060 (38,905; T00156974001)

Figure S1G

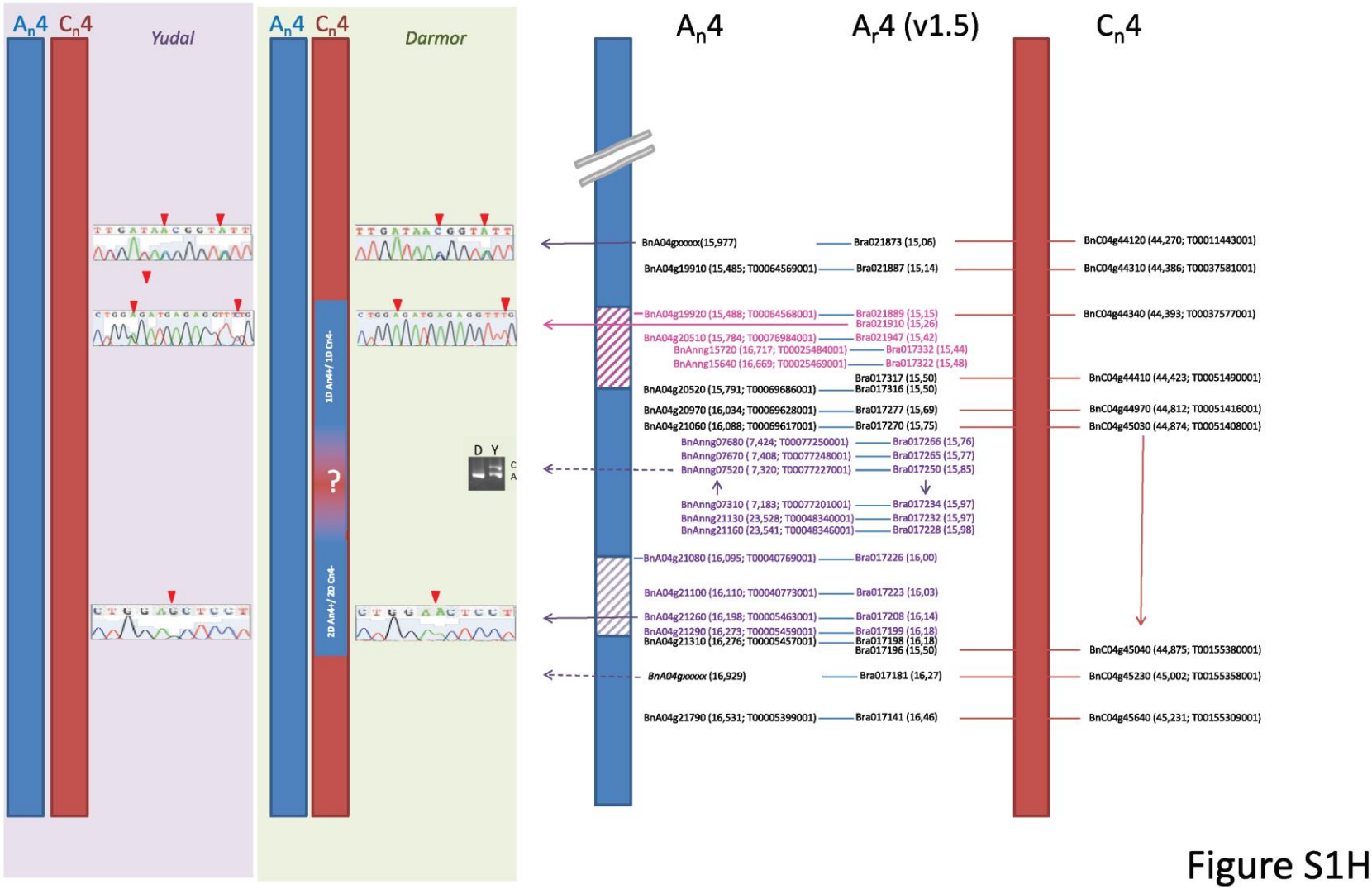

An4 Cn4
Yudal
An4 Cn4
Darmor
1D An4+/ 1D Cn4-
?
2D An4+/ 2D Cn4-
D Y
An4
Ar4 (v1.5)
Cn4
BnA04gxxxxx(15,977)
Bra021873 (15,06)
BnC04g44120 (44,270; T00011443001)
BnA04g19910 (15,485; T00064569001)
Bra021887 (15,14)
BnC04g44310 (44,386; T00037581001)
BnA04g19920 (15,488; T00064568001)
Bra021889 (15,15)
Bra021910 (15,26)
BnC04g44340 (44,393; T00037577001)
BnA04g20510 (15,784; T00076984001)
Bra021947 (15,42)
BnAnng15720 (16,717; T00025484001)
Bra017332 (15,44)
BnAnng15640 (16,669; T00025469001)
Bra017322 (15,48)
Bra017317 (15,50)
BnC04g44410 (44,423; T00051490001)
BnA04g20520 (15,791; T00069686001)
Bra017316 (15,50)
BnA04g20970 (16,034; T00069628001)
Bra017277 (15,69)
BnC04g44970 (44,812; T00051416001)
BnA04g21060 (16,088; T00069617001)
Bra017270 (15,75)
BnC04g45030 (44,874; T00051408001)
BnAnng07680 (7,424; T00077250001)
Bra017266 (15,76)
BnAnng07670 (7,408; T00077248001)
Bra017265 (15,77)
BnAnng07520 (7,320; T00077227001)
Bra017250 (15,85)
BnAnng07310 (7,183; T00077201001)
Bra017234 (15,97)
BnAnng21130 (23,528; T00048340001)
Bra017232 (15,97)
BnAnng21160 (23,541; T00048346001)
Bra017228 (15,98)
BnA04g21080 (16,095; T00040769001)
Bra017226 (16,00)
BnA04g21100 (16,110; T00040773001)
Bra017223 (16,03)
BnA04g21260 (16,198; T00005463001)
Bra017208 (16,14)
BnA04g21290 (16,273; T00005459001)
Bra017199 (16,18)
BnA04g21310 (16,276; T00005457001)
Bra017198 (16,18)
Bra017196 (15,50)
BnC04g45040 (44,875; T00155380001)
BnA04gxxxxx (16,929)
Bra017181 (16,27)
BnC04g45230 (45,002; T00155358001)
BnA04g21790 (16,531; T00005399001)
Bra017141 (16,46)
BnC04g45640 (45,231; T00155309001)

Figure S1H

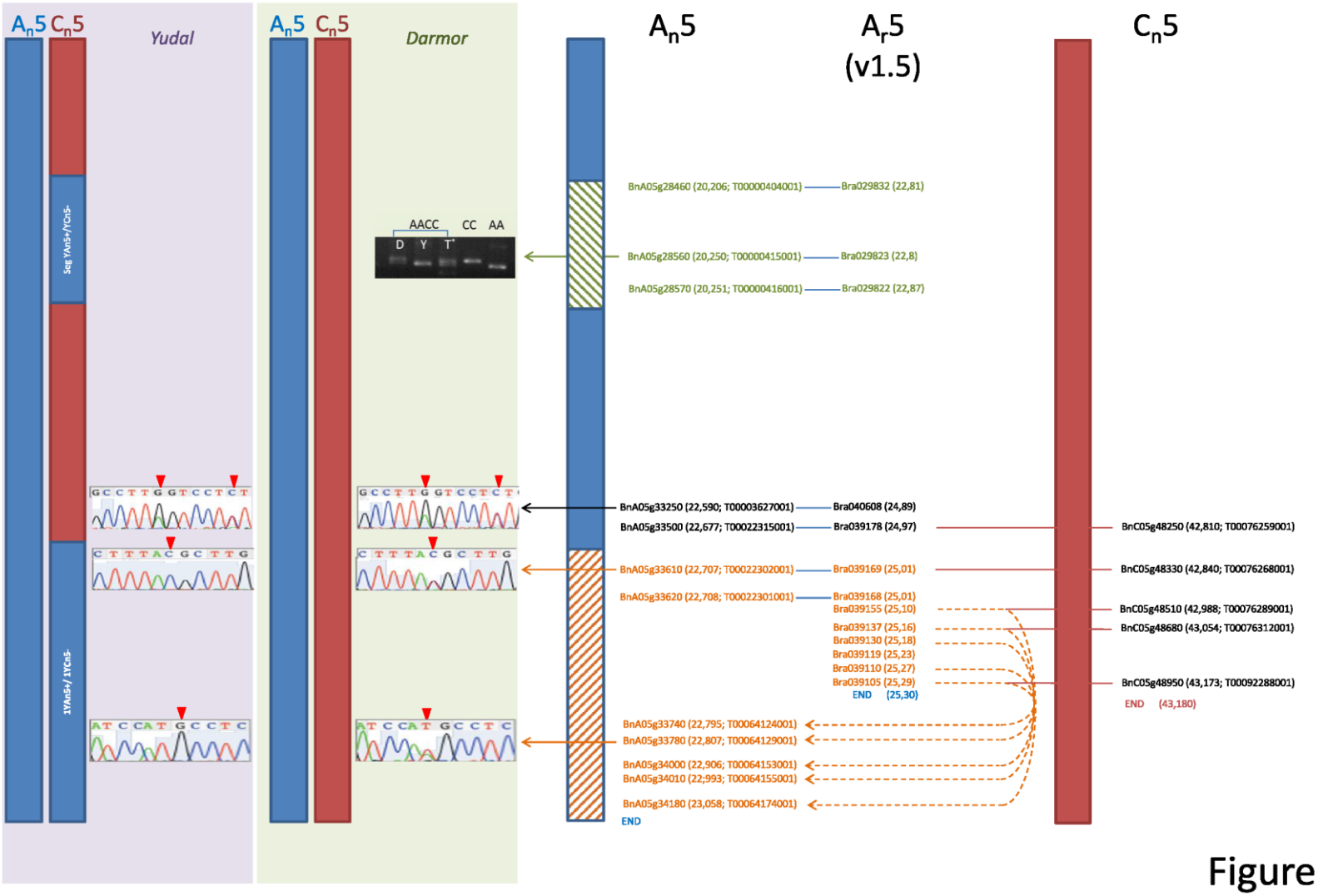

A$_n$5 C$_n$5
Yudal
A$_n$5 C$_n$5
Darmor
A$_n$5
A$_r$5
(v1.5)
C$_n$5
Seq YAn5+/YCn5-
1YAn5+/ 1YCn5-
AACC
CC
AA
D
Y
T
BnA05g28460 (20,206; T00000404001)
Bra029832 (22,81)
BnA05g28560 (20,250; T00000415001)
Bra029823 (22,8)
BnA05g28570 (20,251; T00000416001)
Bra029822 (22,87)
BnA05g33250 (22,590; T00003627001)
Bra040608 (24,89)
BnA05g33500 (22,677; T00022315001)
Bra039178 (24,97)
BnA05g33610 (22,707; T00022302001)
Bra039169 (25,01)
BnA05g33620 (22,708; T00022301001)
Bra039168 (25,01)
Bra039155 (25,10)
Bra039137 (25,16)
Bra039130 (25,18)
Bra039119 (25,23)
Bra039110 (25,27)
Bra039105 (25,29)
END (25,30)
BnA05g33740 (22,795; T00064124001)
BnA05g33780 (22,807; T00064129001)
BnA05g34000 (22,906; T00064153001)
BnA05g34010 (22,993; T00064155001)
BnA05g34180 (23,058; T00064174001)
END
BnC05g48250 (42,810; T00076259001)
BnC05g48330 (42,840; T00076268001)
BnC05g48510 (42,988; T00076289001)
BnC05g48680 (43,054; T00076312001)
BnC05g48950 (43,173; T00092288001)
END (43,180)

Figure S1I

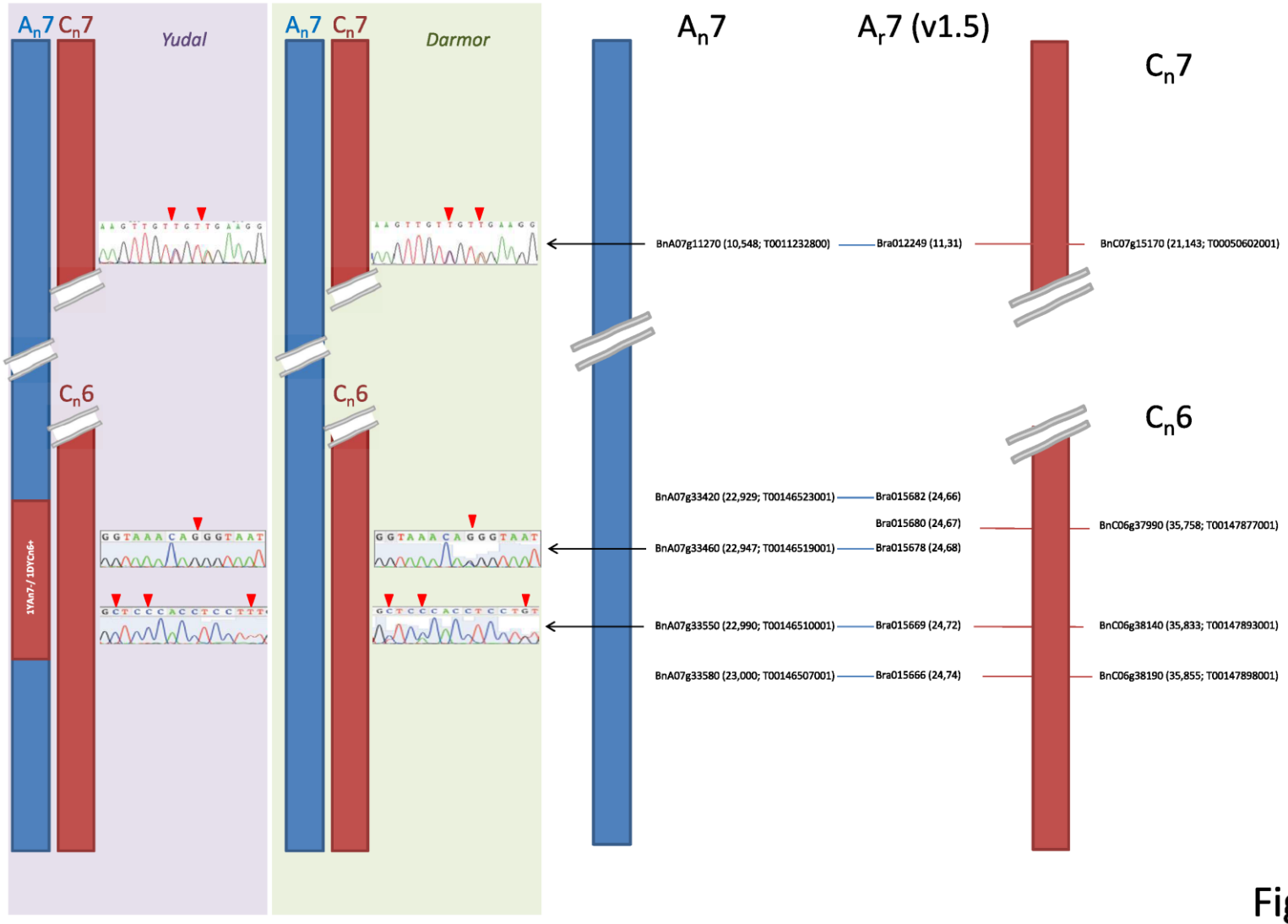

An7 Cn7
Yudal
An7 Cn7
Darmor
An7
Ar7 (v1.5)
Cn7
Cn6
Cn6
1YAn7/1DYCn6
BnA07g11270 (10,548; T0011232800)
Bra012249 (11,31)
BnC07g15170 (21,143; T00050602001)
BnA07g33420 (22,929; T00146523001)
Bra015682 (24,66)
Bra015680 (24,67)
BnC06g37990 (35,758; T00147877001)
BnA07g33460 (22,947; T00146519001)
Bra015678 (24,68)
BnA07g33550 (22,990; T00146510001)
Bra015669 (24,72)
BnC06g38140 (35,833; T00147893001)
BnA07g33580 (23,000; T00146507001)
Bra015666 (24,74)
BnC06g38190 (35,855; T00147898001)

Figure S1J

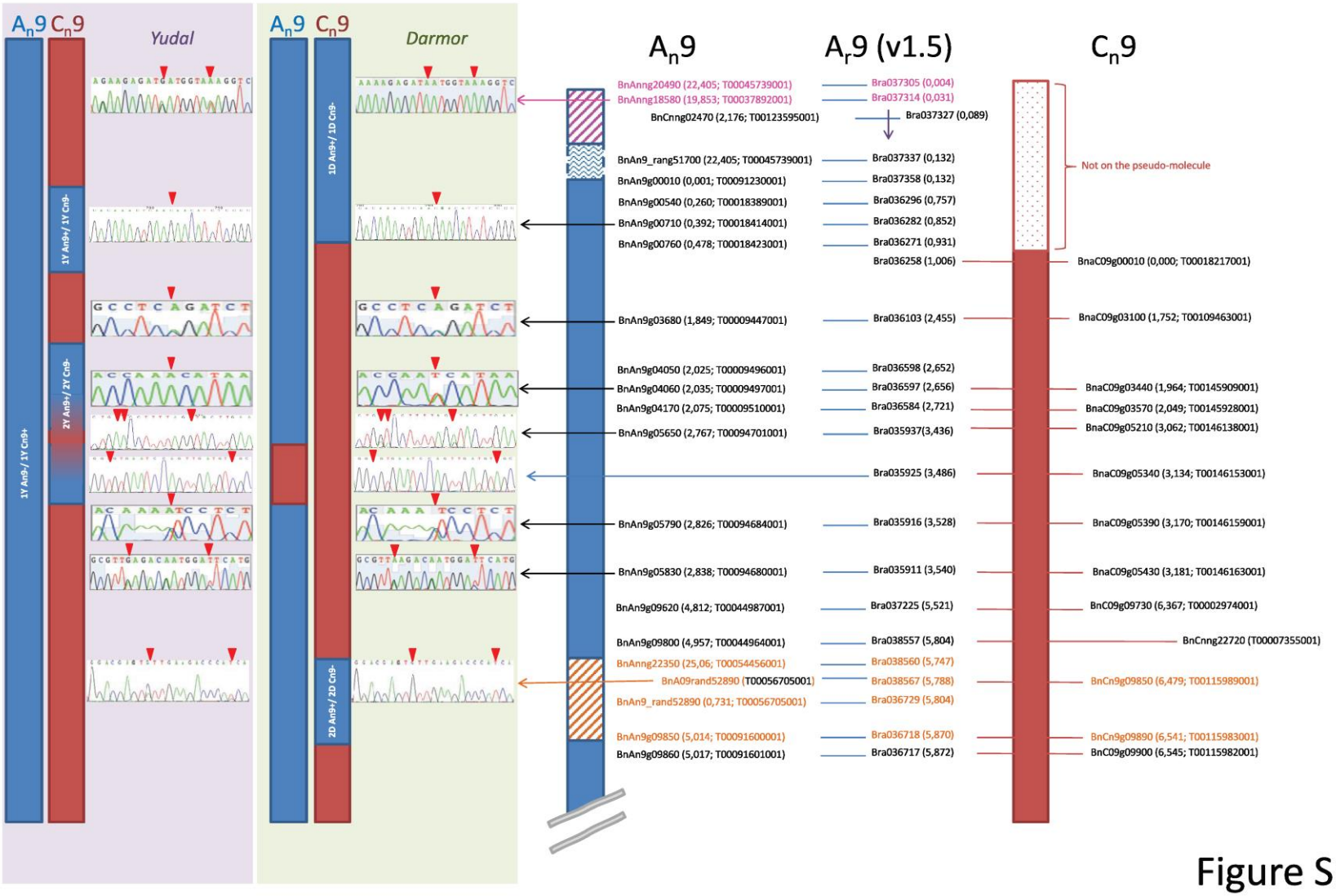

An9 Cn9
Yudal
An9 Cn9
Darmor
An9
Ar9 (v1.5)
Cn9
1Y An9-/ 1Y Cn9+
1Y An9+/ 1Y Cn9-
2Y An9+/ 2Y Cn9-
1D An9+/ 1D Cn9-
2D An9+/ 2D Cn9-
BnAnng20490 (22,405; T00045739001)
BnAnng18580 (19,853; T00037892001)
BnCnng02470 (2,176; T00123595001)
BnAn9_rang51700 (22,405; T00045739001)
BnAn9g00010 (0,001; T00091230001)
BnAn9g00540 (0,260; T00018389001)
BnAn9g00710 (0,392; T00018414001)
BnAn9g00760 (0,478; T00018423001)
BnAn9g03680 (1,849; T00009447001)
BnAn9g04050 (2,025; T00009496001)
BnAn9g04060 (2,035; T00009497001)
BnAn9g04170 (2,075; T00009510001)
BnAn9g05650 (2,767; T00094701001)
BnAn9g05790 (2,826; T00094684001)
BnAn9g05830 (2,838; T00094680001)
BnAn9g09620 (4,812; T00044987001)
BnAn9g09800 (4,957; T00044964001)
BnAnng22350 (25,06; T00054456001)
BnA09rand52890 (T00056705001)
BnAn9_rand52890 (0,731; T00056705001)
BnAn9g09850 (5,014; T00091600001)
BnAn9g09860 (5,017; T00091601001)
Bra037305 (0,004)
Bra037314 (0,031)
Bra037327 (0,089)
Bra037337 (0,132)
Bra037358 (0,132)
Bra036296 (0,757)
Bra036282 (0,852)
Bra036271 (0,931)
Bra036258 (1,006)
Bra036103 (2,455)
Bra036598 (2,652)
Bra036597 (2,656)
Bra036584 (2,721)
Bra035937(3,436)
Bra035925 (3,486)
Bra035916 (3,528)
Bra035911 (3,540)
Bra037225 (5,521)
Bra038557 (5,804)
Bra038560 (5,747)
Bra038567 (5,788)
Bra036729 (5,804)
Bra036718 (5,870)
Bra036717 (5,872)
Not on the pseudo-molecule
BnaC09g00010 (0,000; T00018217001)
BnaC09g03100 (1,752; T00109463001)
BnaC09g03440 (1,964; T00145909001)
BnaC09g03570 (2,049; T00145928001)
BnaC09g05210 (3,062; T00146138001)
BnaC09g05340 (3,134; T00146153001)
BnaC09g05390 (3,170; T00146159001)
BnaC09g05430 (3,181; T00146163001)
BnC09g09730 (6,367; T00002974001)
BnCnng22720 (T00007355001)
BnCn9g09850 (6,479; T00115989001)
BnCn9g09890 (6,541; T00115983001)
BnC09g09900 (6,545; T00115982001)

Figure S1K

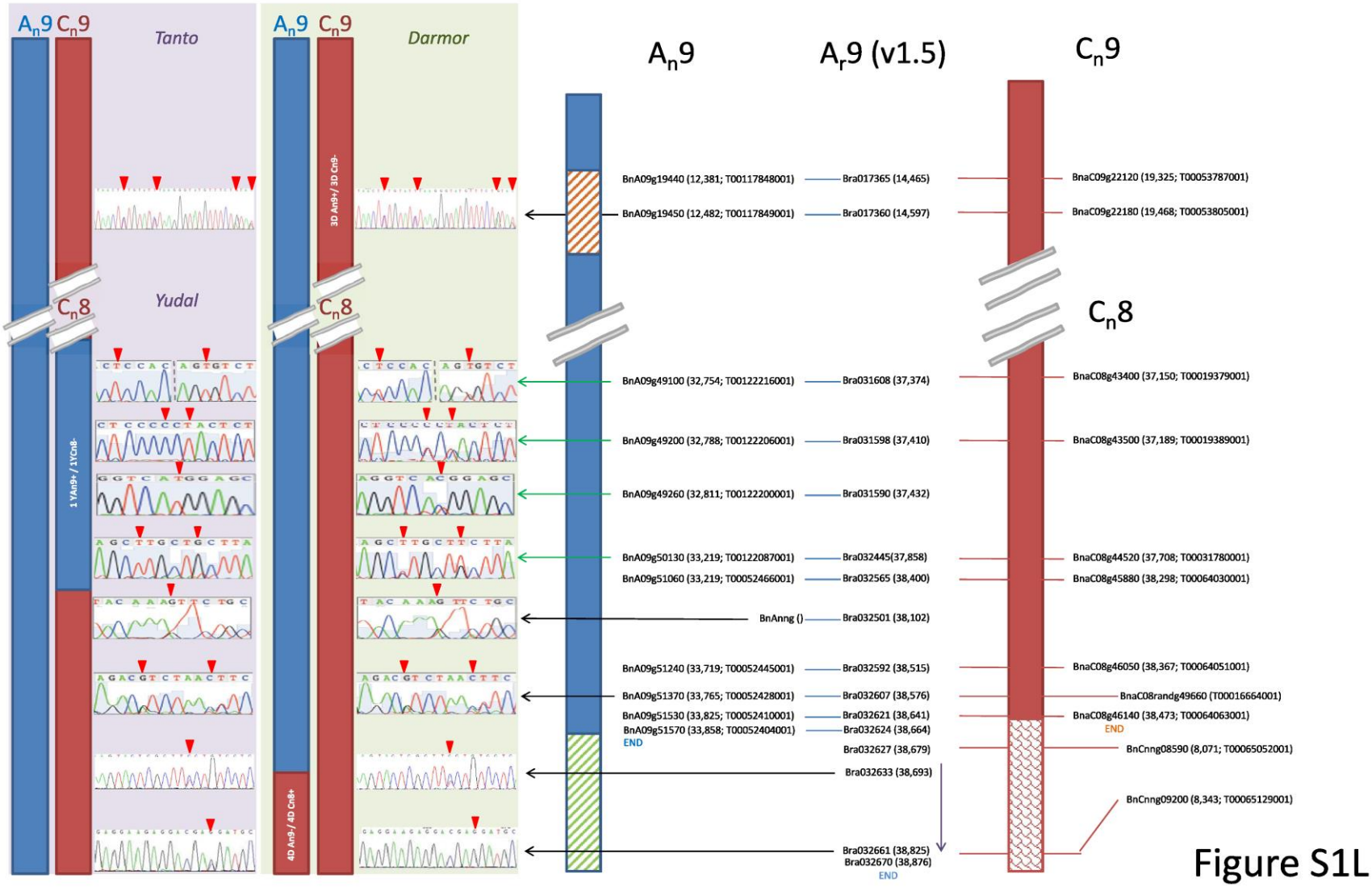

An9 Cn9
Tanto
Yudal
Cn8
1 YAn9+ / 1YCn8-
An9 Cn9
Darmor
3D An9+/ 3D Cn9-
Cn8
4D An9-/ 4D Cn8+
An9
Ar9 (v1.5)
Cn9
BnA09g19440 (12,381; T00117848001)
Bra017365 (14,465)
BnA09g19450 (12,482; T00117849001)
Bra017360 (14,597)
BnaC09g22120 (19,325; T00053787001)
BnaC09g22180 (19,468; T00053805001)
Cn8
BnA09g49100 (32,754; T00122216001)
Bra031608 (37,374)
BnaC08g43400 (37,150; T00019379001)
BnA09g49200 (32,788; T00122206001)
Bra031598 (37,410)
BnaC08g43500 (37,189; T00019389001)
BnA09g49260 (32,811; T00122200001)
Bra031590 (37,432)
BnA09g50130 (33,219; T00122087001)
Bra032445(37,858)
BnaC08g44520 (37,708; T00031780001)
BnA09g51060 (33,219; T00052466001)
Bra032565 (38,400)
BnaC08g45880 (38,298; T00064030001)
BnAnng ()
Bra032501 (38,102)
BnA09g51240 (33,719; T00052445001)
Bra032592 (38,515)
BnaC08g46050 (38,367; T00064051001)
BnA09g51370 (33,765; T00052428001)
Bra032607 (38,576)
BnaC08randg49660 (T00016664001)
BnA09g51530 (33,825; T00052410001)
Bra032621 (38,641)
BnaC08g46140 (38,473; T00064063001)
END
BnA09g51570 (33,858; T00052404001)
Bra032624 (38,664)
END
Bra032679 (38,679)
BnCnng08590 (8,071; T00065052001)
Bra032633 (38,693)
BnCnng09200 (8,343; T00065129001)
Bra032661 (38,825)
Bra032670 (38,876)
END

Figure S1L

Figure S1: Validation of fixed Homoeologous Exchanges in Darmor-bzh and Yudal

On the left are represented the An (blue) and Cn (red) chromosomes in Yudal (purple background) and Darmor-bzh (green background). On the right are listed sorne gene models from the corresponding regions, with their position along the *B. napus* pseudomolecules (in Mb) and their alias (GSBRNA2-) given between brackets. When available, the name (position) of the orthologues in *B. rapa* (Ar;v1.5) is aIso given (homoeologues/orthologues are connected by a line).

The HEs identified by Chalhoub et al. (2014) are represented by the hatched areas superimposed over the An pseudomolecule (on the right); the genes located within an HE are written with the same color as the hatching representing this HE. Genes in black are outside HEs.

Arrows indicate the genes that we used to design primers pairs; every arrow point towards (part of) the chromatogram profiles we obtained by sequencing (Sanger) the corresponding amplicons in *Darmor-bzh* and *Yudal*, respectively. The chromatograms are centered on the SNP(s) that we used to discriminate homoeologous copies (red arrow on top of the chromatogram). ln sorne occasions, when the A and C amplicons differed in size, the chromatograms are replaced a gel (with A/AA and C/CC indicating the size of the band in *B. rapa* and *B. oleracea*, respectively).

The HEs that we validated are indicated by a rectangle superimposed over the An or Cn chromosomes (on the left); blue rectangle superimposed over the Cn red chromosome, HE resulting in the loss of C copies and the concurrent duplication of A copies; red rectangle superimposed over the An blue chromosome, HE resulting in the loss of A copies and the concurrent duplication of C copies.

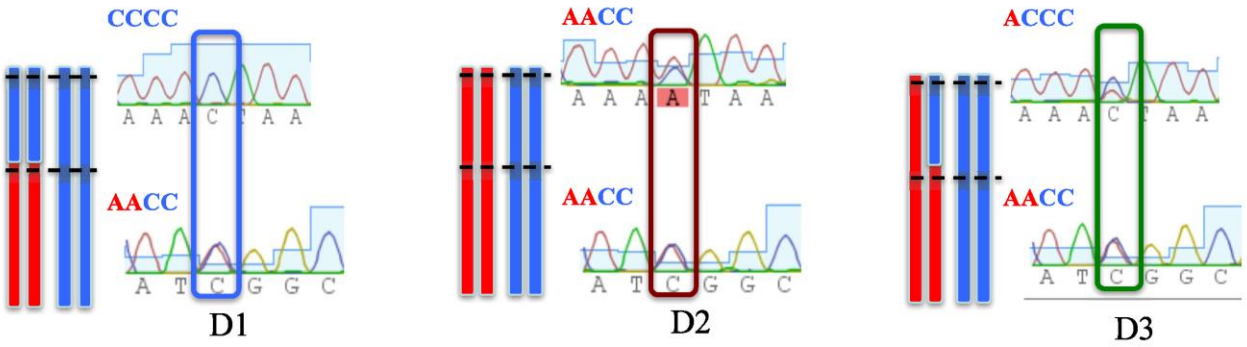


Figure S2: Detection of a *de-novo* Homoeologous exchange segregating between biological replicates in *Darmor-bzh*

For each of the three biological replicates (D1, D2, D3), we represent on the left our expectation in term of A (chromosome red) and C (chromosome blue) copies number based on the expression data. Chromatogram profile corresponding to amplification performed within (top) or outside (bottom) of the expected HE are centered on a SNP discriminating the A from the C copies.

Figure S2

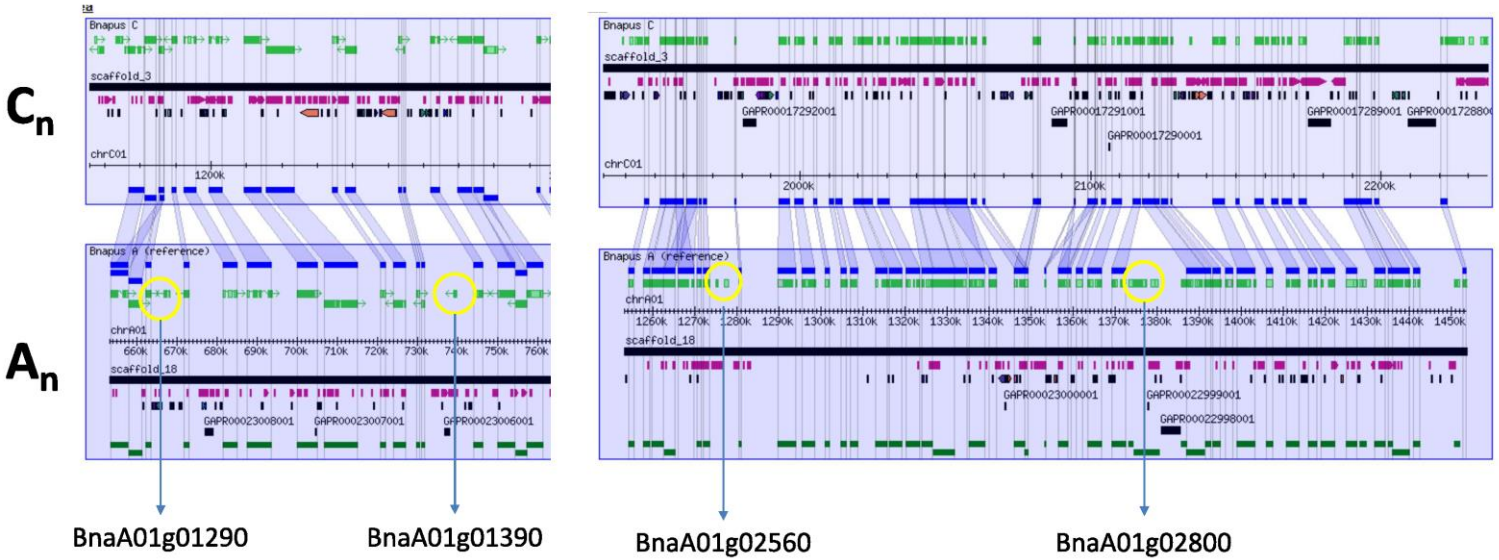


Figure S3: **Evidence for the presence of A-genome specific genes within the region of the newly formed HE.** Here are provided outputs from syntenic comparisons between the homoeologous A (bottom) and C (top) regions that were lost (A) and duplicated (C) following the newly-formed HE. Gene models are represented by colored arrows. Pairs of homoeologues are represented by thick blue lines connected by shaded boxes. The A-genome specific genes, which have no homoeologues in the C region are surrounded by a yellow circle with the gene Ids shown below. We did not count genes for which the missing homoeologues may have been an artefact of inter contig gaps (contigs are represented by thick black lines).

Figure S3

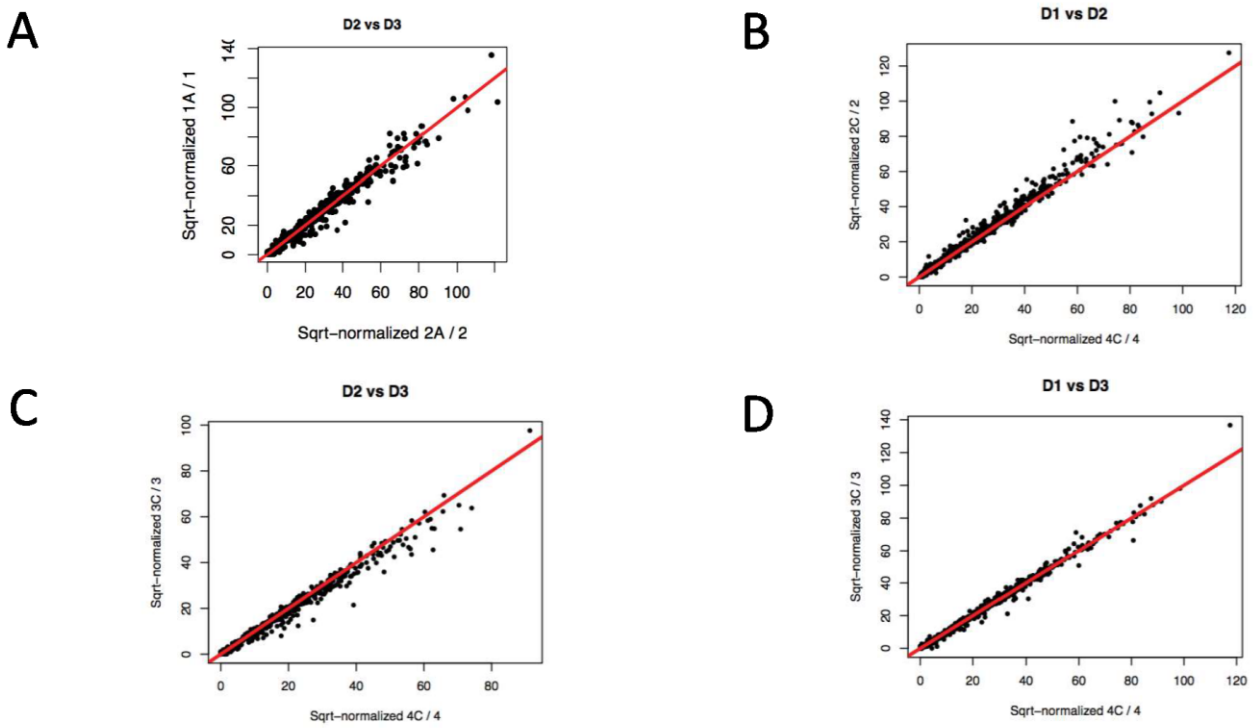


Figure S4: Gene expression is proportional to gene copy number for the newly formed HE.
For genotype pairs, comparisons of gene expression (square-root normalised) divided by gene copy number were made . A genome expression: **A)** AACC (D2; 2A/2) vs ACCC (D3; A/1). C genome expression: **B)** CCCC (D1; 4C/4) vs AACC (D2; 2C/2) **C)** AACC (D2; 2C/2) vs ACCC (D3; 3C/3) **D)** CCCC (D1; 4C/4) vs ACCC (D3; 3C/3). Red line (x = y), expectation if gene expression is in strict concordance with gene copy number.

Figure S4

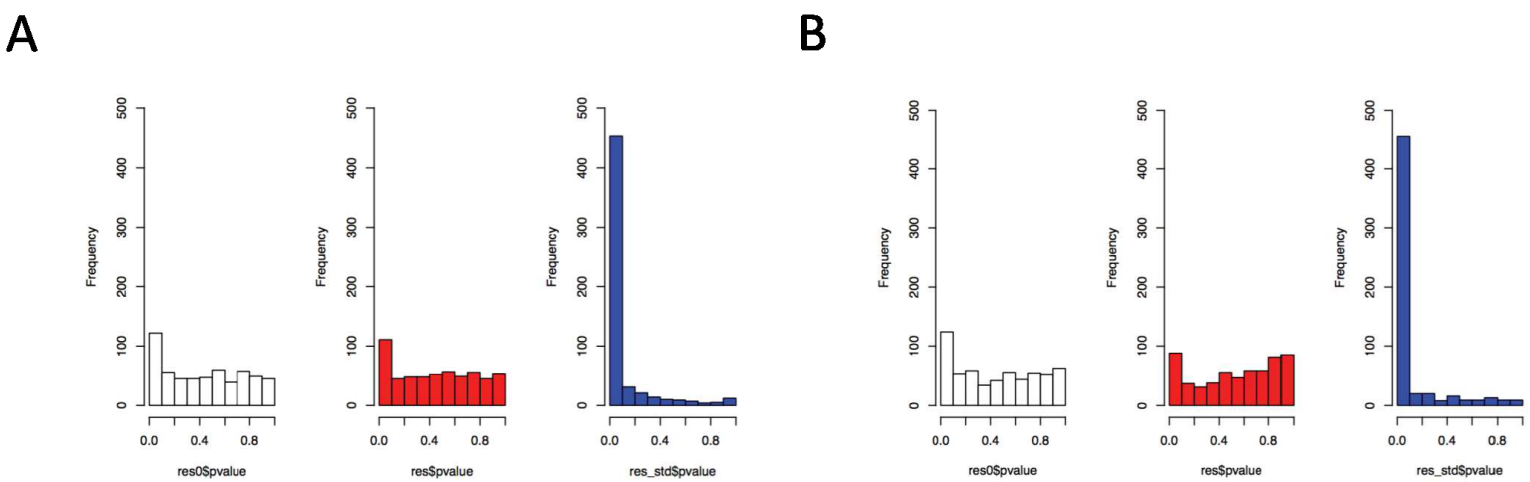


Figure S5: P-value distributions for genes within the newly formed HE.
A) Genes on the A genome, **B)** genes on the C genome. When copy number is not taken into consideration, most genes are differentially expressed between individuals ($p < 0.05$; blue). When expression is normalised by gene copy number, the vast majority of genes are not differentially expressed ($p > 0.05$; red) i.e. gene expression per copy is constant. This is also true when normalising based on the number of reads that map within the HE ($p > 0.05$; white), this last normalisation approximates the gene dosage normalisation (red), which is expected if read mapping is proportional to gene dosage.

Figure S5

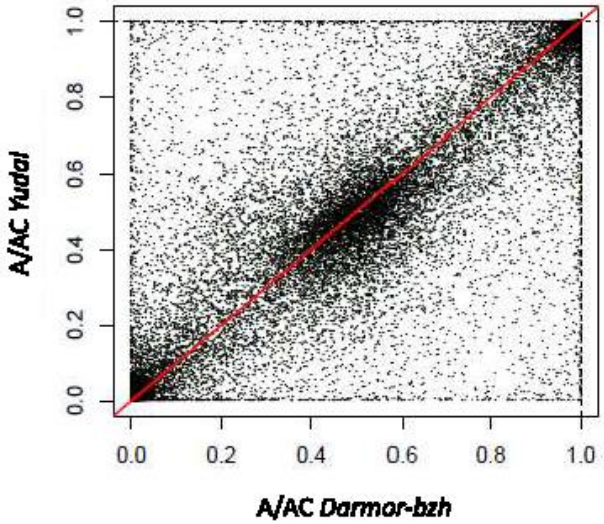


Figure S6: Homeologue bias (contribution of the A genome copy to Total(A+C) expression) in *Darmor-bzh* can be used to predict homoelogue bias in *Yudal*. Red line - A/AC *Darmor-bzh* = A/AC *Yudal*. This plot was drawn using genes that are not located within known HEs.

Figure S6

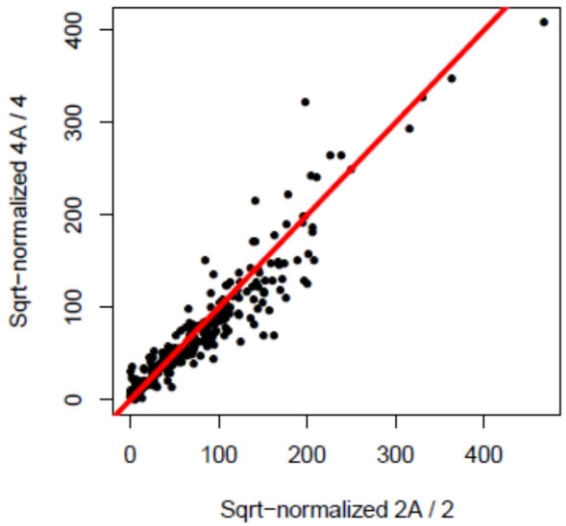


Figure S7: Gene expression is proportional to gene copy number for the fixed HEs in *Yudal*.
Comparison of gene expression (square-root normalised) divided by gene copy number for *Yudal* (AAAA; 4A/4) and *Darmor-bzh* (AACC; 2A/2). Red line (x = y), expectation if gene expression is in strict concordance with gene copy number. Only *Yudal*-specific AACC -> AAAA HEs (which represent the majority) are considered.

Figure S7

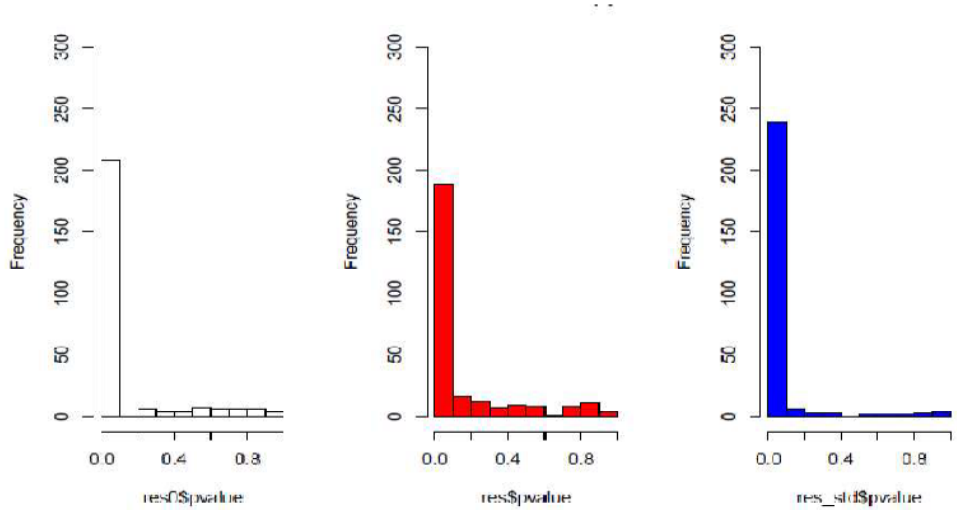


Figure S8: P-value distributions comparing *Darmor-bzh* and *Yudal* expression, for genes within fixed *Yudal*-specific HEs.

When copy number is not taken into consideration, most genes are differentially expressed between *Darmor-bzh* and *Yudal* ($p < 0.05$; blue). When expression is normalised by gene copy number, fewer genes are differentially expressed ($p > 0.05$; red) i.e. copy number contributes to gene expression, however a significant number of genes are still differentially expressed indicating additional inter-varietal contributions to the difference in expression. This is also true when normalising based on the number of reads that map within the HE ($p > 0.05$; white).

Figure S8

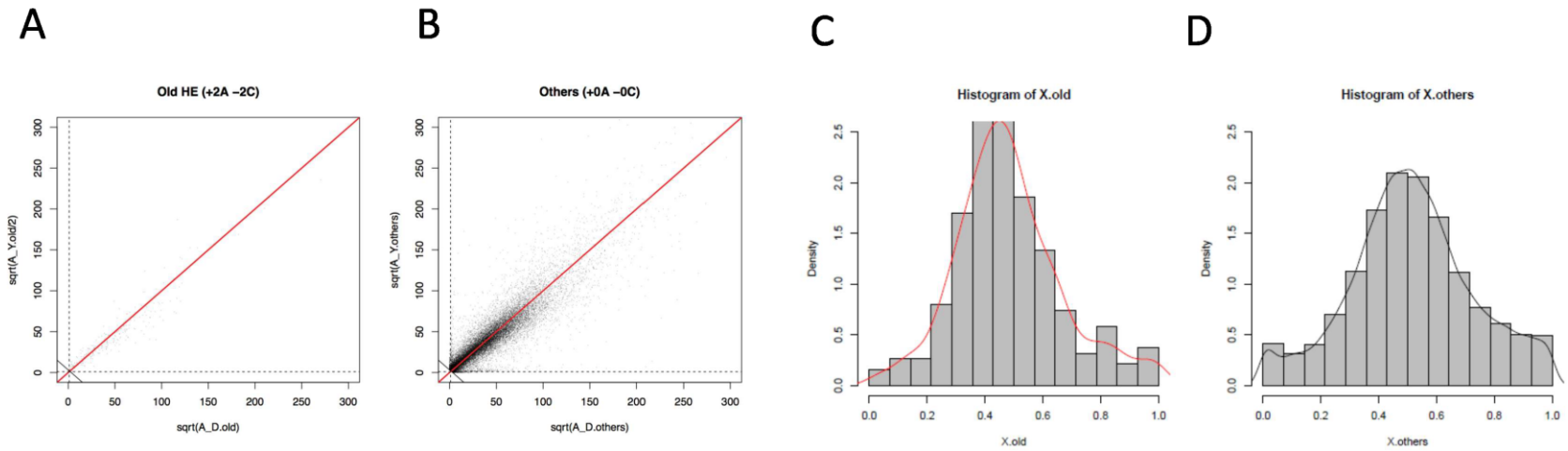


Figure S9: Concordance in expression per copy between *Darmor-bzh* and *Yudal* for genes within *Yudal*-specific HEs and outside HEs.

The square-root of the expression of A genes in *Darmor-bzh* is plotted against the square-root of the expression per copy of A genes in *Yudal* for genes within fixed HEs **(A)** and outside HEs **(B)**. The expected position for genes with equal expression per copy in *Darmor-bzh* and *Yudal* is given by the red lines.

Histograms summerize the difference in expression per copy between *Darmor-bzh* and *Yudal* for genes inside and outside HEs **(C & D** respectively**)**. Specifically, these histograms show the distributions of the angles created for each gene by a line running from the origin (x=0,y=0) to the gene (x=*Darmor-bzh*, y=*Yudal*) in the scatterplots **(A & B).** The angle is divided by p /2 so that the final metric ranges between 0 and 1. Within HEs (X.old): arctangent((*Yudal*/2)/*Darmor-bzh*)/(p /2); outside HEs (X.others) = arctangent(*Yudal*/*Darmor-bzh*)/(p /2).

When *Yudal* and *Darmor-bzh* expressions per copy are equal x = y (i.e. red lines in **A & B**), hence we obtain a value of 0.5 [angle/(p /2)]. The two distributions (**C & D**, red and black lines respectively) differ significantly (two-sample Kolmogorov-Smirnov test, p = 3.2E-4) confirming divergent transcriptional outcomes for genes within HEs.

Figure S9

Supplementary Tables

Table S1: PCR assay to test for the expected loss of one of the two homeologous copies associated with every HE described in Chalhoub et al., 2014

chro	HE event	exp. dose	Validation	Bna A region (Mb)	orthologous Bra region	homeologous Bna C region	target	Forward Primer	Reverse Primer	Tm	size	consensus between		additionnal copies to be excluded				
A01-C01	1DAn1+/1DCn1-	2A - 0C	✓	BnA_random (partially found)	BrA01:26.33-26.64	not found	Bra021434	AGGGTACTCGAGATTTTGCTCA	AGCTCTCCACCCTGCAATTA	59	258	BnA01g02910	Bol001717					
A01-C01	1DAn1+/1DCn1-	2A - 0C	✓	BnA_random (partially found)	BrA01:26.33-26.64	not found	Bra021461	GACCCGAGATGCTCACAAGA	AACAGGAGAGCAGCGGTAAG	59	190	BnaA01g32990D	Bol000730					
A01-C01																		
A01-C01	3DAn1+/3DCn1-	2A - 0C	✓	BnA_random (partially found)	BrA01: 24.39-24.57	not found	Bra038698	TTGCGTAGAGAACTAAGGG	GACACCATATCGCTTACTGA	52	849	BnaA01g30850D	C01:36168043..361	BnaA03g32120D	BnaC03g05720D			
A01-C01	3DAn1+/3DCn1-	2A - 0C	✓	BnA_random (partially found)	BrA01: 24.39-24.57	not found	Bra038686	ATCCATTTAGAACCGCATTA	TGAAGACGGAACCCATTA	50	503	ChUn_random: 112	Bol007273					
A01-C01		2A - 2C	✓				Bra034176	TGGCATAAGTTCGGAGAC	ACACACCCGTTAAATCCTAA	51	970	BnaA01g30870D	C01:36625740..36627739					
A01-C01	4DAn1+/4DCn1-	2A - 0C		BnA01: 21.06-21.52	BrA01: 24.60-24.92	BnC01: 37.63-37.63												
A01-C01		2A - 2C	✓				Bra034081	TCTTCTGACATCCTGATTCTCCT	GCTAGGCAAGAAGAGAGTCC		166							
A01-C01	5DAn1+/5DCn1-	2A - 0C	✓	BnA01: 21.71-21.73	BrA01: 25.10-25.23	BnC01: 37.85-37.96	Bra034074	AAGAGCGGTTAGATAGTTTT	ACTGGAGGCTGTTTTGCTAC	50	418	BnaA01g32200D	C01_37472977..374	A5_2697519	A3_291586	C3_447114	A10_344557	
	5DAn1+/5DCn1-	2A - 2C	X	BnA01: 21.71-21.73	BrA01: 25.10-25.23	BnC01: 37.85-37.96	Bra034072	GAGATGGAACACGGAGACCT	AACTCCCCGAAGTACCCTCT	56	1697	BnaA01g32220D	Bol023085	A5_20391456	C5_40289055	A3_1484490(	C3_21762497	C4_2689191(A5_3174857
A01-C01		2A - 2C	✓				Bra040063	TTTTTGAACGCCTAGAATCA	GACCCCAAATGTAGCAGAC	54	642	BnaA01g32230D	BnaC01g39280D	BnaC03g36300D				
A01-C01	6DAn1+/6DCn1-	2A - 0C	✓	BnA01: 21.81-22.61	BrA01: 25.45-25.89	BnC01: 38.11-38.34	Bra040237	GAAGGCCAACTCTGCCCAT	ATCCACCGTCTGCATCATGT	59	243	chrUn_random:6383351-6383593						
A01-C01	6DAn1+/6DCn1-	2A - 0C	✓	BnA01: 21.81-22.61	BrA01: 25.45-25.89	BnC01: 38.11-38.34	Bra040254	AGAGGATCTTCACTCACCTGT	TGAAGGATGAGCGGCTTCA	59	356	chrUn_random:6319642-6319997						
A01-C01	7DAn1+/7DCn1-	0A - 2C	✓	not found	BrA01: 26.70-26.78	BnaC_random	Bra040517	ACTCGGAAAATCGTAAAATC	CTCGGCTAAGTTCTTCGT	52	774	chrUn_random_84(	Bol158910	BnaA05g38860D	BnaC05g48820D			
A01-C01	7DAn1+/7DCn1-	0A - 2C	✓	not found	BrA01: 26.70-26.78	BnaC_random	Bra040506	ATCCATCACATCGTCCA	TCTTTGTCACCTATACACGC	52	682	chrUn_random_85(	Bol158780	BnaA03g27540D	BnaC03g52550D			
A02-C02	1DAn2+/1DCn2-	2A - 0C	✓	BnaA_random	BrA02: 0.06-0.14	not found	Bra028900	TAGCAGACCCACCGAATG	CACGAGAGAGTGCGACTTG	53	566	BnaAnng00220D	BnaC02g03610D					
A02-C02	1DAn2+/1DCn2-	2A - 0C	✓	BnaA_random	BrA02: 0.06-0.14	not found	Bra028881	TTTACGATGAACCGGATTTC	GCCTCCATTCCCATAAGAT	54	238	BnaAnng00260D	BnaC02g03020D	BnaA03g00960D	BnaC03g01310D			
A02-C02	1DAn2+/1DCn2-	2A - 0C	✓	BnaA_random	BrA02: 0.06-0.14	not found	Bra028629	TTTCTCGGCTGCTCGGAGTT	CATGTGCGGATTCGTGATGC	56	764	BnaA02g00020D	BnaC02g00810D	BnaC03g01310D				
A02-C02	2DAn2+/2DCn2-	2A - 2C	X				Bra020112	ATTTCTTGCTGTATGCGTCT	TGCCACATCTCTCGAACT	53	965	BnaA02g04950D	BnC2g021940	BnaC09g37250D	BnaA10g14880D			
A02-C02	2DAn2+/2DCn2-	2A - 2C	X				Bra020131	ATTTGGGTTATTATTGGTCA	CATGTTTGAACGTCGTGT	51	808	Bra020131	BnC2g023200					
A02-C02	2DAn2+/2DCn2-	2A - 0C	✓	BnA02: 2.03-2.51	BrA02: 3.36-3.87	BnC02: 5.07-5.15	Bra020154	GCAAACACATCCCGGAGTTA	TTAGACGAAGCTGAGGTGGC	59	295	BnaA02g05310D	BnC2g023500					
A02-C02		2A - 2C	✓				Bra020183	TCCTTCCTCTCTTGGCTATG	GCGGACACTTGTTACCAAC	54	605	BnaA02g05550D	BnaC02g09440D					
A02-C02	3DAn2+/3DCn2-	2A - 0C	✓	BnA02: 2.77-3.26	BrA02: 4.60-5.00	not found	Bra020231	ATCTACACTCTCCGCACTTC	TGTTCTGATGTTGCGTATG	52	901	BnaA02g06110D	BnC2g024630	BnaA10g13560D	BnaC09g559(	BnaC03g74690D		
A02-C02	3DAn2+/3DCn2-	2A - 0C	✓	BnA02: 2.77-3.26	BrA02: 4.60-5.00	not found	Bra020285	CAGCGATGAAGGGTATGGAG	CTTGTTCCCTGCGTTCAAA	56	768	BnaA02g06800D	BnC2g025380					
A02-C02		2A - 2C	✓				Bra020314	AAGGAAGAAGGTCGTCGTGA	CAATGCCCACGCCTAAAG	53	939	BnaA02g07030D	BnaC02g09920D	BnaC09g34830D	BnaC03g709(	BnaA08g00190D		
A02-C02	4DAn2-/4DCn2+	0A - 2C	X	BnA02: 10.07-10.22	BrA02: 12.41-12.55	BnC02: 19.84-19.92	Bra008153	TCCTTTGCCTACACCTTTA	GCCTCTTTGTCTTGTCATC	52	678	Bra008153	BnaC02g22910D	BnaA07g81600D	BnaC06g35420D			
A02-C02																		
A02-C02	5DAn2+/5DCn2-	2A - 0C	✓	BnA02: 23.69-24	BrA02: 25.34-25.75	not found	Bra029288	TCGCCGGGAAAGTATAAACT	TTTGGATGAAGCGGTTTAAC	52	735	BnaA02g33160D	BnC2g161860					
A02-C02	5DAn2+/5DCn2-	2A - 0C	✓	BnA02: 23.69-24	BrA02: 25.34-25.75	not found	Bra029256	GGGCAATAGAGTCGAGGAC	CATATTGCCATAACGGAATC	51	548	BnaA02g33420D	BnC2g164180	BnaA06g21930D				
A04-C04		2A - 2C	✓				Bra021873	AGTTATGTTCAGGCGAGTTT	GGCTCATCAGAGTTTCCA	54	1036	Bra021873	BnaC04g44120D					
A04-C04	1DAn4+/1DCn4-	2A - 0C	✓	BnA04: 15.49-15.79	BrA04: 15.15-15.49	BnC04: 44.39-44.42	Bra021910	ACAAAGAAAAACAGAACCGA	GCATCTTATTGGCTTGACC	54	555	Bra021910	Bol037864					
A04-C04																		
A04-C04	2DAn4+/2DCn4-	2A - 0C	✓	BnaA_random	BrA04: 15.76-16.18	not found	Bra017208	CGAAACCTGGACCGCCTTC	ATCGCCGTTGAGTCGTTCCT	57	786	Bra017208	Bol037686					
A09-C09	1DAn9+/1DCn9-	2A - 0C	✓	BnA_random (partially found)	BrA09: 0.004-0.13	BnaC_random	Bra037314	GCATAACCAAGTCACTGCACT	TGTGCTTTCTCTAGGTCTGGG	59	203	chrUn_random:60377403-60377605						
A09-C09																		
A09-C09	2DAn9+/9DCn9-	2A - 0C	✓	BnaA09_random	BrA09: 5.74-5.87	BnaC09: 19.37-19.47	Bra038567	CCGGAGACCAGACAATCT	CAGTCTTTCTAACGCAGCC	56	897	Bra038567	BnaC09g09850D	BnaA06g34660D	BnaA09rg52890D			
A09-C09																		
A09-C09	3DAn9-/3DCn9+	0A - 2C	X	BnaA09: 12.39-12.49	BrA09: 14.46-14.53	BnaC09: 6.46-6.55	Bra017360	TGCTTTCCAGGTTGTTGTC	AGGAGTGATTAAGCGATATGC	53	864	BnaA09g19450D	BnaC09g22180D	BnaC07g22680D	BnaA06g335(	BnaC02g353(	ChrRand_ 75652565	
A09-C08	4DAn9-/4DCn8+	0A - 2C	X	not found	BrA09: 38.67-38.87	BnaC_random	Bra032633	CATATTCGCCGCCATTGT	AACGGGTTCTCACCTTCGTA	56	624	Bra032633	Bo8g117830	BnaC05g00650D	BnaA10g00580D			
A09-C08	4DAn9-/4DCn8+	0A - 2C	✓	not found	BrA09: 38.67-38.87	BnaC_random	Bra032661	CAACACGATGAAGGCAATA	TCTAGCGCCTGAGATCTTAC	56	581	Bra032661	BnaC08random_Br	BnaC04_3727927	BnaA04_264(	BnaC05_431(	BnaA10_17353847	
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra040237	GAAGGCCAACTCTGCCCAT	ATCCACCGTCTGCATCATGT	59	243	chrUn_random:6383351-6383593						
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra040254	AGAGGATCTTCACTCACCTGT	TGAAGGATGACGAGGCTTCA	59	356	chrUn_random:6319642-6319997						
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra021434	AGGGTACTCGAGATTTTGCTCA	AGCTCTCCACCCTGCAATTA	59	258	BnA01g02910	Bol001717					
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra021461	GACCCGAGATGCTCACAAGA	AACAGGAGAGCAGCGGTAAG	59	190	BnaA01g32990D	Bol000730					
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra021469	GCCAGCATCGAGGTTTGTTT	ACGCGTACAAAGACACGAAC	59	171	BnaA01g33060D	Bol000738					
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra040119	CCGTCGGGTATTATTCTCTCTG	TTGCCTCCCTTTCTTCCCTT	59	126	BnaC01g39890D	BnaA01g33400D					
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra040136	GCCGCGACTTACATCATAGATC	TTGTTCTCTCTGCTCTGCCT	59	173	BnaC01g40070D	BnaA01g33560D					
A01-C01	1YAn1-/1YCn1+	0A - 2C	✓	BnA01: 21,72-end (22.41-22.57)	BrA01: 25.26-end (26.64-26.65)	BnaC01: 37,97-end	Bra040151	TGTCTCGCATGTAAGGAGCA	GCAACAGCACCAAAACCATT	59	153	BnaC01g40230D	BnaA01g33700D					
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra028900	TAGCAGACCCACCGAATG	CACGAGAGAGTGCGACTTG	53	566	BnaAnng00220D	BnaC02g03610D					
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra028881	TTTACGATGAACCGGATTTC	GCCTCCATTCCCATAAGAT	54	238	BnaAnng00260D	BnaC02g03020D	BnaA03g00960D	BnaC03g01310D			
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra028629	TTTCTCGGCTGCTCGGAGTT	CATGTGCGGATTCGTGATGC	56	764	BnaA02g00020D	BnaC02g00810D	BnaC03g01310D				
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra028560	AACCTGGTGATGATGGCAAC	GCAGGCTCGTAATCATGTGG	59	216	BnaA02g00750D	BnaC02g00030D					
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra028552	TCGTCTTCATCCTCAGCGTT	ATGGATCGGTGGTTCGTAGG	59	167	BnaA02g00830D	chrUn_random:30525427-30525593					
A02-C02	1YAn2+/1YCn2-	2A - 0C	✓	BnA02: 0-0.42	BrA02: 0-1.8 (1.63-1.75)	BnaC02: 0-1.8	Bra023312	TCAGTTCAAAGCTCGACACAC	GCCCACTGTTTATGGTGTAGG	59	151	BnaA02g00880D	chrUn_random:30559272-30559422					
A02-C02		2A - 2C	✓				Bra023331	AGCACTCTCCAGCATCTCTC	CCGACGAATGAGCTGAAACA	59	246	BnaA02g01070D	BnaC02g03920D					
A02-C02																		
A02-C02	2YAn2+/2YCn2-	2A - 0C	X	BnA02: 2.09-2.51	BrA02: 3.95-5.38	not assembled	Bra020154	GCAAACACATCCCGGAGTTA	TTAGACGAAGCTGAGGTGGC	59	295	BnaA02g05310D	BnC2g023500					
A02-C02																		
A02-C02	3YAn2+/3YCn2-	2A - 0C	✓	BnA02: 4.55-4.85	BrA02: 6.51-6.80	BnaC02:8.5-8.88	Bra035553	CCGCAGATGAGGGTCAATGT	GCCTATGACGGGAACGACTG	56	727	BnaA02g09280D	BnaC02g13350D	chrC09_34768281				
A02-C02	3YAn2+/3YCn2-	2A - 0C	✓	BnA02: 4.55-4.85	BrA02: 6.51-6.80	BnaC02: 8.5-8.88	Bra022737	TCTCCTTGGTTTGGTGGTT	TCTGGGAAGAACAAGTGGTC	52	1472	BnaA02g09510D	BnaC02g13570D	BnaC09g53250D	BnaA10g288(	BnaA01g162(	BnaC01g19340D	
A02-C02																		
A02-C02	4YAn2+/4YCn2-	2A - 0C	✓	BnA02: 24.14-24.27	BrA02:25.87-26.02	BnaC02: 45.07-45.21	Bra029232	CGAGTGGAAATGCGGTTAC	CATCTTACGACGGCTTTGTT	53	1190	BnaA02g33680D	BnaC02g42440D	BnaC09g14260D				
A03-C03	1YAn3+/1YCn3-	2A - 0C	✓	BnA03: 1.32-1.47	BrA03: 1.73-1.85	BnaC03: 1.97-2.1	Bra006022	CATAAGCTCGAGACGCCTTC	AGCTGGACTCATCGCTGG	59	171	BnaA03g55840D	BnaC03g03810D					
A03-C03	1YAn3+/1YCn3-	2A - 0C	X	BnA03: 1.32-1.47	BrA03: 1.73-1.85	BnaC03: 1.97-2.1	Bra006027	AGTTTGATCACTGCCTCCCT	GGACAGCTTCGTCTGGAGAT	59	161	BnaA03g55950D	BnaC03g03860D					
A03-C03	1YAn3+/1YCn3-	2A - 0C	✓	BnA03: 1.32-1.47	BrA03: 1.73-1.85	BnaC03: 1.97-2.1	Bra006035	AGGCGTCGATGTCTGGAC	CAACTTCTCCTCCGCCTCC	59	167	BnaA03g55870D	BnaC03g03980D					
A03-C03	1YAn3+/1YCn3-	2A - 0C	✓	BnA03: 1.32-1.47	BrA03: 1.73-1.85	BnaC03: 1.97-2.1	Bra006052	TGCTGGAGTCGACAAAGTGA	CCATCAAGAACTCTCCTTCGC	59	159	BnaA03g02830D	BnaC03g04180D					

A03-C03	1YAn3+/1YCn3-	2A - 0C	✓	BnA03: 1.32-1.47	BraA03: 1.73-1.85	BnaC03: 1.97-2.1	Bra006068	GTCTCTTAATCTTGCTCGCCA	GAGGAAAGGGAAGTTACCGAAG	59	272	BnaA03g03020D	BnaC03g04400D			
A03-C03		2A- 2C	✓				Bra006073	GGATCTCCTGCACGCATATG	TGTGATAAAGCTCCCTGATCTGA	59	206	BnaA03g03080D	BnaC03g04500D			
A03-C03		2A - 2C	✓				Bra006081	CTCTTTGCTGCCTACCCCT	AGGGTGATGAAATTCTGGAGAGA	59	178	BnaA03g03150D	BnaC03g04560D			
A05-C05		2A - 2C	✓				Bra040608	TGAAGAGCTGCTCCAAGAGG	ACGATGGGAGCAGAGGATTT	59	158	BnaA05g33250D	chrUn_random:8653058-8653215			
A05-C05	1YAn5+/1YCn5-	2A - 0C	✓	BnA05: 22.66-23.07	BraA05: 24.97-end	BnaC05: 42.8-43.18	Bra039169	TTGGGGACTAGGAAGCATGG	AACCAGACTGAGGACCAAGG	59	215	BnaA05g33610D	BnaC05g48330D			
A05-C05	1YAn5+/1YCn5-	2A - 0C	✓	BnA05: 22.66-23.07	BraA05: 24.97-end	BnaC05: 42.8-43.18	Bra039110	ACCTCATCAAAACCGTCTTCG	AGAGTGATGGTGGCTGTGAA	59	173	BnaA05g33780D	BnaC05g48900D			
A07-C06	1YAn7-/1YCn6+	0A - 2C	✓	BnA07: 22.93-23	BrA07: 24.66-24.74	BnC06: 35.74-35.84	Bra015678	GATGCTGTTTGTTGCGTTGA	GTGATGCGGAGGTCTACCA	55	1080	BnaA07g33460D	BnaC06g38010D	BnaA02g17930D	ChrUn_Bra015678	
A07-C06	1YAn7-/1YCn6+	0A - 2C	✓	BnA07: 22.93-23	BrA07: 24.66-24.74	BnC06: 35.74-35.84	Bra015669	AACCTTTCTTGGGATCTGG	AGGAAGCGTGGTCTGAGTAG	54	726	BnaA07g33550D	BnaC06g38140D	BnaC02g13700D	BnaC08g08860D	
A09-C08	1YAn8+/1YCn8-	2A - 0C	✓	BnA09: 32.87-33.69	BraA09:37.48-38.48	BnaC08: 37.21-38.34	Bra031608	ATGGGATAAACGTCGGGCTT	AAGACCAGCCACACAGAGG	59	199	BnaA09g49100D	BnaC08g43400D			
A09-C08	1YAn8+/1YCn8-	2A - 0C	✓	BnA09: 32.87-33.69	BraA09:37.48-38.48	BnaC08: 37.21-38.34	Bra031598	TGGAACGATGAGGTTTGATGC	AACCTCAACAGCCACAACAC	59	151	BnaA09g49190D	BnaC08g43490D			
A09-C08	1YAn8+/1YCn8-	2A - 0C	✓	BnA09: 32.87-33.69	BraA09:37.48-38.48	BnaC08: 37.21-38.34	Bra031590	AACCCTAGCACTCCCTCTCT	AAAGCCAAGATCGAACGAGC	59	160	BnaA09g49260D	chrUn_random:56786219-56787436			
A09-C08	1YAn8+/1YCn8-	2A - 0C	✓	BnA09: 32.87-33.69	BraA09:37.48-38.48	BnaC08: 37.21-38.34	Bra032445	CAGATCTGACTCAGCTGCAT	TTGTTTCTGCAGTCTCGTCT	59	246	BnaA09g50130D	BnaC08g44520D			
A09-C08		2A - 2C	✓				Bra032501	AGAAGCTGCATAGTCCACCA	GGGAGAGAACATACCCAAGAAG	59	150	chrUn_random:362	BnaC08g50180D			
A09-C09	1YAn8+/1YCn8-	2A - 0C	✓	BnA09: 0.26-0.48	BraA09: 0.74-0.93	BnaC_random	Bra036282	CGGTAACTACAGCGATGGAA	AGTCTGAGCCGTTATCCAAC	53	980	BnaA09g00710D	BnaCnng23260D			
A09-C09		2A - 2C	✓				Bra036103	GCAACGATGGACTCAGCTTT	GGAGAGAGTACACAAGTTTGCA	59	160	BnaA09g03680D	BnaC09g03100D			
A09-C09	2YAn8+/2YCn8-	2A - 0C	✓	BnA09: 1.13-2.07	BraA09: 1.69-2.45	BnaC09: 0.91-2.05	Bra036597	TGGCGATTCTCAGTAGCCAA	GGAATCTGTAAGCGTAGTCAGC	59	234	BnaA09g04060D	BnaC09g03440D			
A09-C09		2A - 2C	✓				Bra035916	TCAAATTGGCAGCAGTCTAGC	TGGAATCATAGACGAAGAGCAAG	59	164	BnaA09g05780D	BnaC09g05380D			
A09-C09		2A - 2C	✓				Bra035911	AGAAGAGCTTGAAGGAGTTAAGC	GCATTCCGTCTCGCATCTG	59	189	BnaA09g05830D	BnaC09g05430D			
A09-C09	1YAn8-/1YCn8+	2A - 0C	✗	BnA09: 2.75-2.78	BrA09: 3.42-3.49	BnC09: 3.02-3.14	Bra035925	ACACTGGTGAGAAAGGGTAT	TTGCGTCTTTGATTTCG	54	983	BnaAnng26590D	BnaC09g05340D	BnaA03g40300D	BnaC07g126	BnaC02g42130D

✓ Validated
✓ Unexpected event detected
✗ Not Validated

Table S2: SNP assay to test for the expected loss of one of the two homeologous regions associated with every HE described in Chalhoub et al. 2014 or disclosed by this study

chro	HE event	expectation	Number of SNP markers within the HE [1]	Number of SNP markers validating the HE [2]
A01-C01	1DAn1+/1DCn1-	loss of C	3	0
A01-C01	3DAn1+/3DCn1-	loss of C	0	NA
A01-C01	4DAn1+/4DCn1-	loss of C	8	0
A01-C01	5DAn1+/5DCn1-	loss of C	0	NA
A01-C01	6DAn1+/6DCn1-	loss of C	2	0
A01-C01	7DAn1+/7DCn1-	loss of C	0	NA
A02-C02	1DAn2+/1DCn2-	loss of C	4	0
A02-C02	2DAn2+/2DCn2-	loss of C	2	0
A02-C02	3DAn2+/3DCn2-	loss of C	8	1
A02-C02	4DAn2-/4DCn2+	loss of A	0	NA
A02-C02	5DAn2+/5DCn2-	loss of C	0	NA
A04-C04	1DAn4+/1DCn4-	loss of C	0	NA
A04-C04	2DAn4+/2DCn4-	loss of C	0	NA
A09-C09	1DAn9+/1DCn9-	loss of C	0	NA
A09-C09	2DAn9+/9DCn9-	loss of C	0	NA
A09-C09	3DAn9-/3DCn9+	loss of A	0	NA
A09-C08	4DAn9-/4DCn8+	loss of A	1	0
A01-C01	1YAn1-/1YCn1+	loss of A	4	4
A02-C02	1YAn2+/1YCn2-	loss of C	59	33
A02-C02	2YAn2+/2YCn2-	loss of C	2	NA
A02-C02	3YAn2+/3YCn2-	loss of C	31	16
A02-C02	4YAn2+/4YCn2-	loss of C	0	NA
A03-C03	Seg YAn3-/YCn3+	loss of A	3	2
A03-C03	1YAn3+/1YCn3-	loss of C	5	0
A05-C05	1YAn5+/1YCn5-	loss of C	15	5
A05-C05	Seg YAn5+/YCn5-	loss of C	8	4
A07-C06	1YAn7-/1YCn6+	loss of A	3	2
A09-C08	1YAn9+/1YCn8-	loss of C	21	5
A09-C09	1YAn9+/1YCn9-	loss of C	2	0
A09-C09	2YAn9+/2YCn9-	loss of C	10	0
A09-C09	1YAn9-/1YCn9+	loss of C	0	NA

[1]: inclusion within an HE region was determined based on Best Blast Hit scores and the HE boundaries defined by Chalhoub et al., 2014
Note, no SNPs invalidate HEs as there is no way to decipher whether the SNPs are located within the duplicated or the lost regions associated with every HE

[2]: snp for which the genotype call is "failed" for the variety where the HE leads to a loss

Table S3: Gene counts within HEs

chro	HE event	exp. dose	Bna A region		orthologous Bra region		homeologous Bna C region		orthologous Bol region			
			size (kb)	# gene models	size (kb)	# gene models	size (kb)	# gene models	size (kb)	# gene models		
A01-C01	1DAn1+/1DCn1-	2A - 0C	NA	56	316	67	NA	13		70		
A01-C01	3DAn1+/3DCn1-	2A - 0C	NA	24	175	29	NA	6		36		
A01-C01	4DAn1+/4DCn1-	2A - 0C	464	57	327	59	-	-		87		
A01-C01	5DAn1+/5DCn1-	2A - 0C	20	10	126	26	110	17		37		
A01-C01	6DAn1+/6DCn1-	2A - 0C	823	80	579	117	230	31		144		
A01-C01	7DAn1+/7DCn1-	0A - 2C	NA	-	470	19	NA	19		22		
A02-C02	1DAn2+/1DCn2-	2A - 0C	NA	90	70	91	NA	69		30		
A02-C02	2DAn2+/2DCn2-	2A - 0C	487	90	510	94	NA	17		132		
A02-C02	3DAn2+/3DCn2-	2A - 0C	490	61	400	65	NA	1		92		
A02-C02	4DAn2-/4DCn2+	0A - 2C	150	12	140	20	NA	15		27		
A02-C02	5DAn2+/5DCn2-	2A - 0C	310	44	410	56	NA	5		73		
A04-C04	1DAn4+/1DCn4-	2A - 0C	300	57	340	71	NA	10		94		
A04-C04	2DAn4+/2DCn4-	2A - 0C	NA	52	420	65	NA	5		86		
A09-C09	1DAn9+/1DCn9-	2A - 0C	NA	19	128	28	NA	15		28		
A09-C09	2DAn9+/9DCn9-	2A - 0C	NA	9	81	13	62	6		21		
A09-C09	3DAn9-/3DCn9+	0A - 2C	100	4	70	6	139	6		21		
A09-C08	4DAn9-/4DCn8+	0A - 2C	NA	-	200	43	NA	43		56		
		Total D	**2680**	**665**	**4762**	**869**	**541**	**278**		**1056**	**943**	
A01-C01	1YAn1-/1YCn1+	0A - 2C	160	208	874	225	330	187		238		
A02-C02	1YAn2+/1YCn2-	2A - 0C	NA	365	1753	354	1800	366		388		
A02-C02	2YAn2+/2YCn2-	2A - 0C	407	72	410	76	NA	6		73		
A02-C02	3YAn2+/3YCn2-	2A - 0C	438	75	440	74	NA	90		88		
A02-C02	4YAn2+/4YCn2-	2A - 0C	130	23	145	29	137	28		31		
A03-C03	Seg YAn3-/YCn3+	0A - 2C	110	31	86	32	111	29		29		
A03-C03	1YAn3+/1YCn3-	2A - 0C	165	30	175	40	232	51		48		
A05-C05	1YAn5+/1YCn5-	2A - 0C	410	80	320	80	360	86		86		
A05-C05	Seg YAn5+/YCn5-	2A -0C	NA	13	125	25	140	26		25		
A07-C06	1YAn7-/1YCn6+	0A - 2C	71	17	84	17	107	17		16		
A09-C08	1YAn9+/1YCn8-	2A - 0C	904	212	965	227	1057	232		212		
A09-C09	1YAn9+/1YCn9-	2A - 0C	228	30	183	28	NA	30		29		
A09-C09	2YAn9+/2YCn9-	2A - 0C	940	165	1138	149	1214	151		163		
A09-C09	1YAn9-/1YCn9+	2A - 0C	30	11	71	17	118	18		16		
		Total Y	**3993**	**1332**	**6769**	**1373**	**5606**	**1317**		**1442**	**2649**	
		Total	**6673**	**1997**	**11531**	**2242**	**6147**	**1595**		**2498**	**3592**	4740

Table S4: HEs are enriched in up-regulated (duplicated region) or down-regulated (lost region) genes, respectively

chro	HE event	score[1]	score[1]	P-value A	P-value A	P-value C	P-value C	n.A[4]	n.C[4]
Darmor-bzh		A	C	nA regions[2]	nA random[3]	nC regions[2]	nC random[3]		
A01-C01	1DAn1+/1DCn1-	4,92	NA	0,0232	0,00233	NA	NA	44	0
A01-C01	3DAn1+/3DCn1-	-1,93	NA	0,765	0,789	NA	NA	18	0
A01-C01	4DAn1+/4DCn1-	3,4	NA	0,0356	0,00248	NA	NA	63	0
A01-C01	5DAn1+/5DCn1-	0,0961	NA	0,412	0,413	NA	NA	9	0
A01-C01	6DAn1+/6DCn1-	2,37	NA	0,0357	0,00649	NA	NA	69	0
A01-C01	7DAn1+/7DCn1-	NA	NA	NA	NA	NA	NA	0	0
A02-C02	1DAn2+/1DCn2-	-0,375	NA	0,515	0,515	NA	NA	19	0
A02-C02	2DAn2+/2DCn2-	1,97	NA	0,0571	0,00861	NA	NA	76	0
A02-C02	3DAn2+/3DCn2-	-1,11	NA	0,746	0,791	NA	NA	56	0
A02-C02	4DAn2-/4DCn2+	-0,00497	NA	0,567	0,567	NA	NA	11	0
A02-C02	5DAn2+/5DCn2-	-0,605	NA	0,589	0,608	NA	NA	38	0
A04-C04	1DAn4+/1DCn4-	-0,489	NA	0,548	0,562	NA	NA	34	0
A04-C04	2DAn4+/2DCn4-	1,47	NA	0,108	0,0666	NA	NA	41	0
A09-C09	1DAn9+/1DCn9-	-1,46	NA	0,708	0,729	NA	NA	18	0
A09-C09	2DAn9+/9DCn9-	2,13	NA	0,217	0,196	NA	NA	10	0
A09-C09	3DAn9-/3DCn9+	4,62	13,5	0,749	0,758	0,925	0,946	4	6
A09-C08	4DAn9-/4DCn8+	NA	2,54	NA	NA	0,93	0,977	0	42
Yudal									
A01-C01	1YAn1-/1YCn1+	36,8	-32,2	0,00861	0	3.22e-05	0	67	62
A02-C02	1YAn2+/1YCn2-	-13	36,8	0,00417	0	0,00266	0	69	60
A02-C02	2YAn2+/2YCn2-	-20	36,8	0,00314	0	0,00621	0	36	27
A02-C02	3YAn2+/3YCn2-	NA	NA	NA	NA	NA	NA	0	0
A03-C03	1YAn3+/1YCn3-	-36,8	36,8	0,000756	0	0,00895	0	25	24
A05-C05	1YAn5+/1YCn5-	-29,2	36,8	0,00111	0	0,00859	0	38	37
A07-C06	1YAn7-/1YCn6+	36,8	-36,8	0,011	0,000113	0,00204	3.22e-05	12	11
A09-C09	1YAn9+/1YCn9-	-14,9	23,8	0,00848	0,00124	0,00597	0,00066	15	14
A09-C09	2YAn9+/2YCn9-	-21	36,8	0,00439	1.61e-05	0,012	4.83e-05	19	15
A09-C08	1YAn9+/1YCn8-	-23,6	36,8	0,000161	0	0,00478	0	121	105

[1]score is the median log(p-value) multiplied by the sign of fold change for genes within the HE; +ve = upregulated, -ve = downregulated

[2]P-value n regions: p-value obtained comparing the region score to all other contiguous regions with the same number of genes

[3]P-value n random: p-value obtained comparing the region score to the scores obtained from sets of n randomly selected genes

[4]number of genes within each region

For a more detailed explanation see pg 19 of the supplementary statistics

Table S5: Segmentation of Gene expression along B. napus chromosomes

Segment	Chromosome	Number of gene models	first gene model	last gene model	NRHEs	corresponding segment identified using B. rapa coord.
1	A01	49	BnaA01g33220D (GSBRNA2T00057220001) / Bra021477	BnaA01g34480D (GSBRNA2T00015429001) / Bra021367	1Y An1-/ 1Y Cn1+	1
2	A02	49	BnaA02g00020D (GSBRNA2T00143574001) / Bra028629	BnaA02g00880D (GSBRNA2T00143473001) / Bra023312	1YAn2+/1YCn2-	2
3	A03	15	BnaA03g00050D (GSBRNA2T00049181001) / Bra005660	BnaA03g00300D (GSBRNA2T00049212001) / Bra005680	Seg YAn3- / YCn3+	3
4	A07	12	BnaA07g33380D (GSBRNA2T00146527001) / Bra015686	BnaA07g33550D (GSBRNA2T00146510001) / Bra015669	1YAn7-/1YCn6+	5
5	A09	105	BnaA09g02070D (GSBRNA2T00009882001) / Bra039075	BnaA09g04170D (GSBRNA2T00009510001) / Bra036584	2YAn9+/2YCn9-	6
6	A09	15	BnaA09g05470D (GSBRNA2T00094721001) / Bra035955	BnaA09g05710D (GSBRNA2T00094695001) / Bra035919	1YAn9- / 1YCn9+	7
7	A09	79	BnaA09g49410D (GSBRNA2T00122180001) / Bra031577	BnaA09g50980D (GSBRNA2T00052478001) / Bra032557	1YAn9+/1YCn8-	8
8	C01	31	BnaC01g39370D (GSBRNA2T00018566001) / Bra040054	BnaC01g40440D (GSBRNA2T00044481001) / Bra036430	1Y An1-/ 1Y Cn1+	9
9	C02	88	BnaC02g00010D (GSBRNA2T00069054001) / Bra028558	BnaC02g03740D (GSBRNA2T00088816001) / Bra028893	1YAn2+/1YCn2-	10
10	C02	39	BnaC02g12750D (GSBRNA2T00122671001) / Bra035593	BnaC02g13600D (GSBRNA2T00122796001) / Bra022734	3YAn2+/3YCn2-	11
11	C02	13	BnaC02g42370D (GSBRNA2T00082662001) / Bra029237	BnaC02g42720D (GSBRNA2T00068670001) / Bra029201	4YAn2+/4YCn2-	12
12	Cnn	11	BnaCnng19990D (GSBRNA2T00097753001) / Bra008243	BnaCnng20200D (GSBRNA2T00098289001) / Bra023317	1YAn2+/1YCn2-	10
13	C03	20	BnaC03g03830D (GSBRNA2T00134579001) / Bra006024	BnaC03g04460D (GSBRNA2T00140144001) / Bra006069	1YAn3+/1YCn3-	13
14	C05	46	BnaC05g48280D (GSBRNA2T00076263001) / Bra039176	BnaC05g48940D (GSBRNA2T00092287001) / Bra039105	1YAn5+/1YCn5-	15
15	C08	94	BnaC08g43540D (GSBRNA2T00019393001) / Bra031583	BnaC08g45970D (GSBRNA2T00064042001) / Bra032584	1YAn9+/1YCn8-	16
16	C09	92	BnaC09g01630D (GSBRNA2T00109648001) / Bra039061	BnaC09g03500D (GSBRNA2T00145916001) / Bra036591	2YAn9+/2YCn9-	17

Here, segmentation is based on the gene position along B. napus chromosomes; as a consequence, groups of genes affected by the same HE but located on different "chromosomes" in the assembly (e.g. C02 and Cnn-random for segment 9 and 12) appears as separate segments

Table S6: Segmentation of Gene expression along B. rapa chromosomes

Segment	Chromosome	Number of gene models	first gene model	last gene model	NRHEs	corresponding segment identified using B.napus coord.
1	A01	55	BnaA01g32470D (GSBRNA2T00057490001) / Bra040359	BnaAnng36490D (GSBRNA2T00012104001) / Bra021489	1Y An1-/ 1Y Cn1+	1
2	A02	15	BnaAnng13950D (GSBRNA2T00017501001) / Bra028893	BnaAnng00500D (GSBRNA2T00119585001) / Bra028861	1YAn2+/1YCn2-	2
3	A03	15	BnaA03g00050D (GSBRNA2T00049181001) / Bra005660	BnaA03g00300D (GSBRNA2T00049212001) / Bra005680	Seg YAn3- / YCn3+	3
4	A05	49	BnaA05g33490D (GSBRNA2T00022316001) / Bra039179	BnaA05g33740D (GSBRNA2T00064124001) / Bra039105	1YAn5+/1YCn5-	
5	A07	12	BnaA07g33380D (GSBRNA2T00146527001) / Bra015686	BnaA07g33550D (GSBRNA2T00146510001) / Bra015669	1YAn7-/1YCn6+	4
6	A09	108	BnaA09g02070D (GSBRNA2T00009882001) / Bra039075	BnaA09g04170D (GSBRNA2T00009510001) / Bra036584	2YAn9+/2YCn9-	5
7	A09	16	BnaA09g05470D (GSBRNA2T00094721001) / Bra035955	BnaA09g05710D (GSBRNA2T00094695001) / Bra035919	1YAn9- / 1YCn9+	6
8	A09	112	BnaA09g49390D (GSBRNA2T00122182001) / Bra031579	BnaAnng25810D (GSBRNA2T00068399001) / Bra032626	1YAn9+/1YCn8-	7
9	C01	55	BnaCnng52190D (GSBRNA2T00055673001) / Bra040359	BnaCnng29200D (GSBRNA2T00017368001) / Bra021489	1Y An1-/ 1Y Cn1+	8
10	C02	194	BnaC02g03740D (GSBRNA2T00088816001) / Bra028893	BnaCnng20200D (GSBRNA2T00098289001) / Bra023317	1YAn2+/1YCn2-	9 & 12
11	C02	47	BnaC02g12750D (GSBRNA2T00122671001) / Bra035593	BnaCnng16670D (GSBRNA2T00091425001) / Bra022718	3YAn2+/3YCn2-	10
12	C02	14	BnaC02g42370D (GSBRNA2T00082662001) / Bra029237	BnaC02g42720D (GSBRNA2T00068670001) / Bra029201	4YAn2+/4YCn2-	11
13	C03	21	(GSBRNA2T00134579001) / Bra006024	BnaC03g04460D (GSBRNA2T00140144001) / Bra006069	1YAn3+/1YCn3-	13
14	C05	9	BnaC05g42730D (GSBRNA2T00110930001) / Bra029833	BnaC05g42870D (GSBRNA2T00110909001) / Bra029824	Seg YAn5+/YCn5-	
15	C05	48	BnaC05g48280D (GSBRNA2T00076263001) / Bra039176	BnaC05g48940D (GSBRNA2T00092287001) / Bra039105	1YAn5+/1YCn5-	14
16	C09	17	BnaCnng75920D (GSBRNA2T00023699001) / Bra036311	BnaCnng23260D (GSBRNA2T00007838001) / Bra036282	1YAn9+/1YCn9-	
17	C09	98	BnaC09g01630D (GSBRNA2T00109648001) / Bra039061	BnaC09g03500D (GSBRNA2T00145916001) / Bra036591	2YAn9+/2YCn9-	16
18	C08	98	BnaC08g45440D (GSBRNA2T00031669001) / Bra031593	BnaC08g45970D (GSBRNA2T00064042001) / Bra032584	1YAn9+/1YCn8-	15

Here, segmentation is performed using the position of the B. rapa orthologues to every B. napus gene; by doing so, B. napus genes located on the Ann_random and Cnn_random chromosomes can be clustered together with the other B. napus genes

Table S7: Overview the different segmentation analyses

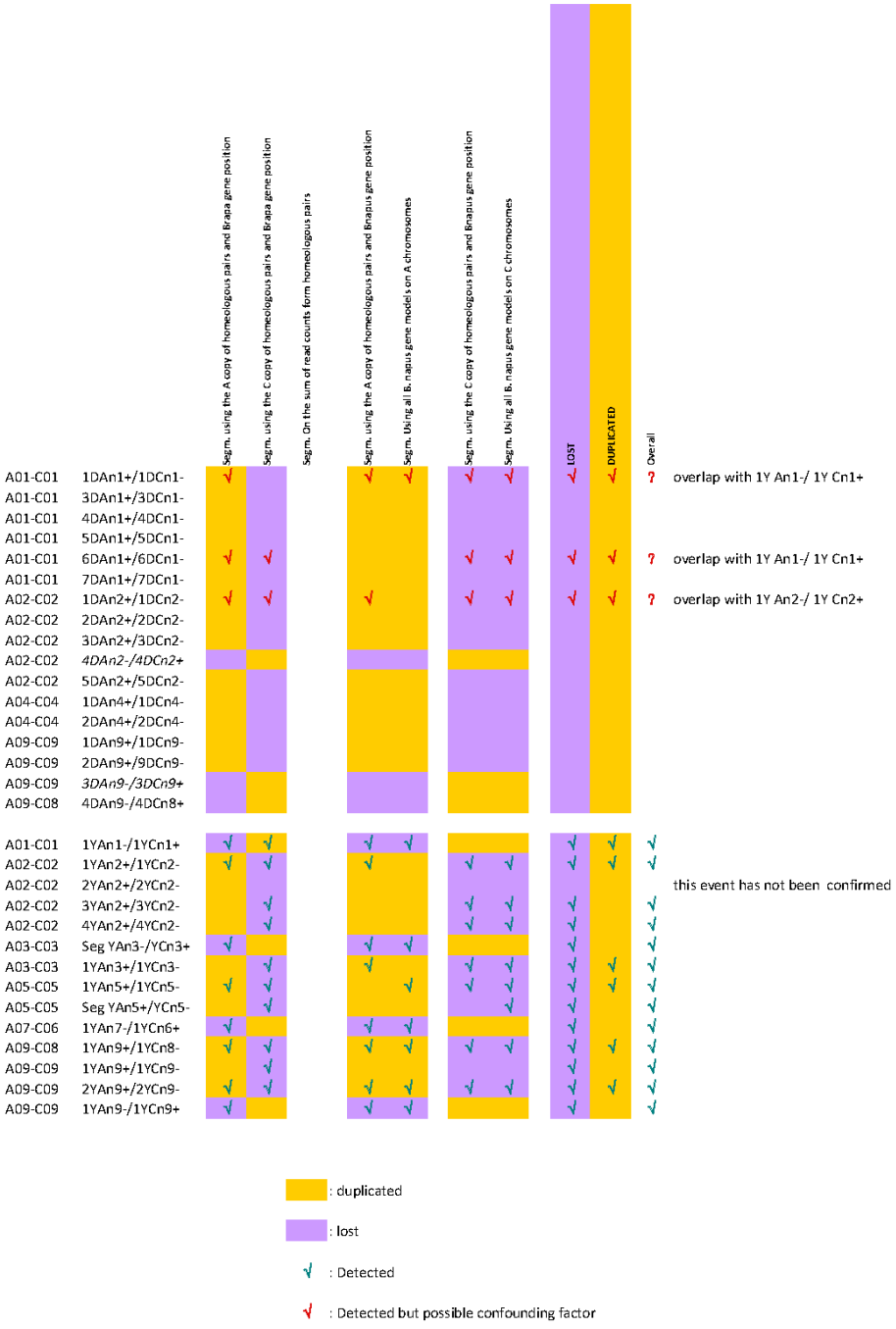


Table S8: PCR assay to test for the expected loss of one of the two homeologous copies associated with the two additionnal HE discloded by this study

chro	HE event	exp. dose	Validation	Bna A region (Mb)	orthologous Bra region	homeologous Bna C region	target	Forward P	Reverse P	Tm	size	consensus between		additionnal copies to be excluded
A03-C03	Seg YAn3-/YCn3+	0A - 2C	√	BnA03: 0-0.10	BraA03: 0.01-0.105	BnaC03: 0-0.123	Bra005667	ACAACTGGGCTTACCGAGAG	CGCCATAACATCCTCGTC	53	728	BnaA03g00160D	BnaC03g00050D	
A03-C03	Seg YAn3-/YCn3+	0A - 2C	√	BnA03: 0-0.10	BraA03: 0.01-0.106	BnaC03: 0-0.124	Bra005672	GAGCATTGGCGTTTGTAACT	CACTCGTTTCTGCGTTCA	54	746	BnaA03g00210D	BnaC03g00100D	
A05-C05	Seg YAn5+/YCn5-	2A -0C	√	BnA05: 22.20-...	BraA05: 22.81-22.92	BnC05: 39.96-40.09	Bra029823	GCAAGAGAACTGGCACGAC	GTTAGAAGAAGGCGACCGA	54	652	BnaA05g28560	BnaC05g42880	

Table S9: Description of RNA-Seq reads obtained by sequencing cDNA with the Illumina technology (paired ends reads) from meiocyes of B. napus var 'Darmor-bzh' and 'Yudal'

Variety	Library designation	Biological replicate	Technical replicate	Total full length read-pairs	Uniquely mapped reads [2]
Darmor-bzh	D11	1	1	30 399 350	55 488 721
Darmor-bzh	D12	1	2	35 177 204	62 982 088
Darmor-bzh	D21	2	1	30 976 273	56 301 493
Darmor-bzh	D22	2	2	33 490 928	60 303 362
Darmor-bzh	D31	3	1	29 829 429	54 259 409
Darmor-bzh	D32	3	2	32 046 080	57 342 388
Yudal	Y11	1	1	35 053 442	57 683 648
Yudal	Y12	1	2	36 364 975	58 732 231
Yudal	Y21	2	1	29 695 891	48 924 269
Yudal	Y22	2	2	36 536 255	58 350 033
Yudal	Y31	3	1	36 993 811	61 239 376
Yudal	Y32	3	2	40 896 029	67 068 253

[1]: number of paired 101-bp long reads

[2]: reads mapped to a unique location on the *Darmor-bzh* genome

Table S10: Expression biases for genes Outside HEs, within fixed *Yudal* HEs, and within the newly formed HE

	Outside HEs (*Yudal* _AACC vs *Darmor-bzh* _AACC) Homoeologue bais prior to HE - (*Darmor-bzh* - A/[A+C])			Fixed *Yudal* HEs (Yudal_AAAA vs Darmor-bzh_AACC) Homoeologue bais prior to HE - (*Darmor-bzh* - A/[A+C])					Newly Formed HE Homoeologue bais prior to HE - (*Darmor-bzh*_ AACC - A/[A+C])				
Yudal $_{(A+C)}$ vs *Darmor-bzh* $_{(A+C)}$	A < C	A ~ C	A > C	A < C	A ~ C	A > C	p-value*	p-value$_{(Others)}$ #	*Darmor-bzh* _CCCC$_{(A+C)}$ vs *Darmor-bzh*_ AACC$_{(A+C)}$	A < C	A ~ C	A > C	p-value*
$Y_{(A+C)} < D_{(A+C)}$	3238	2206	3642	38	19	35	0,73	0,51	$D\_CCCC_{(A+C)} < D\_AACC_{(A+C)}$	119	71	59	6,0E-06
$Y_{(A+C)} \sim D_{(A+C)}$	1630	1107	1043	25	26	18	0,29	0,3	$D\_CCCC_{(A+C)} < D\_AACC_{(A+C)}$	39	45	49	0,29
$Y_{(A+C)} > D_{(A+C)}$	3301	2142	3360	17	24	61	3,6E-07	0,011	$D\_CCCC_{(A+C)} < D\_AACC_{(A+C)}$	46	56	87	3,6E-04
Total	8169	5455	8045	80	69	114	0,015	0,015	Total	204	172	195	0,69

*Binomial test: (A < C) vs (A > C)

#Fisher Exact test: Fixed Yudal HEs vs Outside HEs

0,503823856

Supplementary Material

title: "Analysis of recent event in Euploid Darmor samples"

author: Guillem Rigaill

output: pdf_document

Overview of the data

In total we have 3 Darmor samples and 3 Yudal samples. For each we have 2 technical replicates. Thus, in total we have 12 measurements per gene. We call Di (respectively Yi) the i-th Darmor sample (respectively Yudal sample). We call Dij (respectively Yij) the j-th replicate of Di (respectively Yi).

Pre-processing of the data

For subsequent analysis we kept only genes with more than 10 reads in total. After this filtering step we recovered 63642 genes out of 101040. We then split the data in two: 1. genes identified as within HEs in Chaloub et al. that we will call old HE genes; 2. all other genes (62198 in total).

Identification and analysis of the recent HE.

Identification and characterization of recent HE using segmentation

To identify recent HE genes we ordered the "none old HE genes" along the genome and compared gene expression levels between pairs of Darmor samples. We considered all pairs except direct technical replicates (D11, D12), (D21, D22) and (D31, D32) making a total of 12 pairs. For each considered pair we raised gene-counts to the power 1/4 and then ran the segmentation approach described in Rigaill et al 2012 to detect abrupt changes in the gene expression ratio between this pair of Darmor samples. These changes delimit segments of the genome where the gene expression ratio was particularly far from the expected 1. We classified a gene as part of a newly formed HE if for all 12 pairs it was in a segment with a gene expression ratio higher than 5/4 or lower than 4/5.

Visualization of recent HE identified using segmentation

This way we identified 62198 genes on chromosome A01, A01 random and C01. In the graph below we summed technical replicates (D11+D12, D21+D22,

D31+D32) and represented the log expression ratio between (D1, D2), (D1, D3) and (D2, D3) along chromosome A01, A01 random and C01. Genes classified as part of a newly formed HE are represented in red.

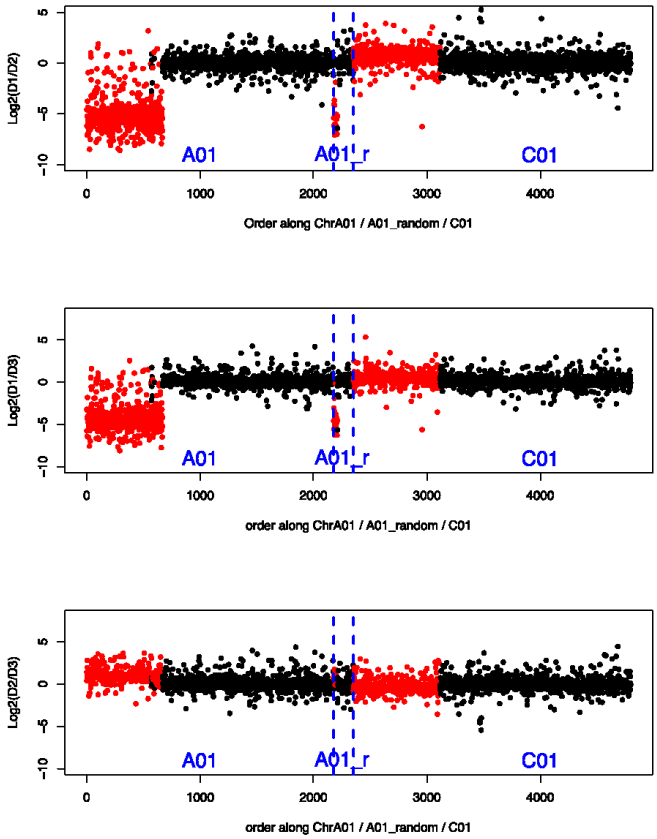


Figure 1: Log-Ratio between the three biological replicates on chromosome A01, A01 random and C01. (top) D1 vs D2, (middle) D1 vs D3, (bottom) D2 vs D3.

Within those 62198 new HE genes we found 571 pairs of homeolog.

Computing the normalization factor on none HE genes

For the following analysis the normalization factor (or size factor) that will be using for the deseq2 analysis is important. We estimated this normalization factor using genes that are not within the old HEs nor within the new HE. We call these genes none HE genes. In total we have 60741 such genes.

Copy number effect on gene expression

In all analyses that follow we will use the 571 A and C homeologs of the newly formed HE to assess whether there is a proportional effect of copy number on gene expression.

A homeolog

To assess whether gene expression for genes in the newly formed HE is proportional to DNA copy number we compared the results of three deseq2 analysis (Deseq2).

In each analysis we compared gene expression between sample D2 with 2 copies of A and sample D3 with one copy of A. The only difference between these 3 analyses is the size factor (normalization factor) used. In details we considered three approaches: 1. The intern approach: in this analysis we use the size factors obtained by deseq2 using the 571 A homeologs of the newly formed HE only. 2. The extern approach: in this analysis we use the size factors obtained by deseq2 on 60741 none HE genes. 3. The Extern*DNAcopy approach: in this analysis we use the size factors obtained by deseq2 on none HE genes and then multiply those by the DNA copy of each sample (2 for sample D2 and 1 for sample D3).

With the Intern approach the size factors should compensate any copy number effect and we expect to detect few differentially expressed genes. With the Extern approach, on the contrary we expect to find many differentially expressed genes. Indeed, although it is not a priori clear that 2 DNA copies lead to a 2 fold gene expression increase, for most genes we expect that 2 DNA copies lead to a higher gene expression than 1 DNA copy. Finally, with the Extern*DNAcopy approach we expect to see many changes if gene expression is not proportional to DNA copy and few if it is proportional.

In deseq2 we used a log-likelihood ratio-test to assess whether gene expression was the same in the two samples (D2 and D3).

The following figure compares for all new HE genes the average A homeolog expression of D2 with 2 copies of A and D3 with 1 copy of A. The data was

normalized using the Extern*DNAcopy approach. It can be seen that data-points are roughly on the $y = x$ line.

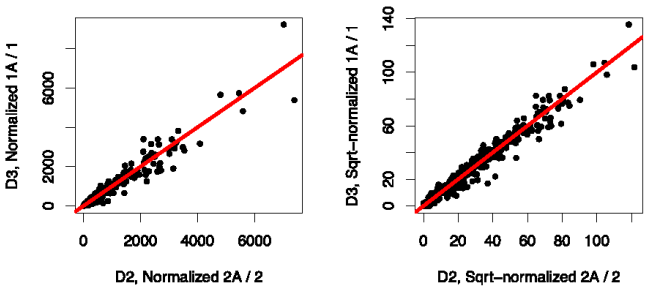


Figure 2: Normalized A count in the newly formed HE genes. Data have been normalized using the deseq2 size factors obtained on genes outside of the old and new HE multiplied by the A copy number (2 for sample D2 and 1 for sample D3). The red-line represent y=x. (Left) Normalized count. (Right) Square-rooted normalized count.

Out of 571 A homeologs the p-value obtained using the Extern*DNAcopy approach is larger than the one obtained with the Extern approach for 511 genes. For the Extern*DNAcopy only 35 genes are found significant with a adjusted p-value smaller than 5 percent compared to 417 for the Extern approach. This difference is highly significant, Fisher's exact test:

```
## [1] 6.020872e-127
```

The following figure represents as histograms the p-values obtained comparing D2 (2A) with D3 (1A) using deseq2 with the Intern (white), Extern*DNAcopy (red) and Extern (blue) normalization approaches.

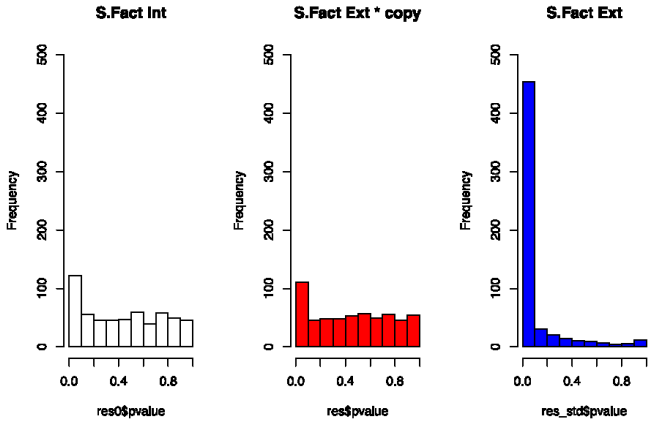


Figure 3: Histograms of p-values obtained using deseq2 on the A homeologs of the new HE genes with either the Intern (white), Extern*DNAcopy (red) and Extern (blue) size factors.

C homeolog

We confirmed this result for the C homeologs of the new HE. For the C homeologs we have 3 samples to consider D1 with 4 copies of C, D2 with 2 copies and D3 with 3 copies. Again we ran deseq2 using a log-likelihood ratio test to assess whether C gene expression was the same in all three samples (D1, D2 and D3) using either the Intern, Extern or Extern*DNAcopy approaches.

The following figure compares for all new HE genes the averaged C homeolog expression of D1 with 4 copies of C, , D2 with 2 copies and D3 with 3. The data was normalized using the Extern*DNAcopy approach. It can be seen that data-points are roughly on the $y = x$ line.

Out of 571 C homeologs the p-value obtained using the Extern*DNAcopy approach is smaller than the one obtained with the Extern approach for 505 genes. With the Extern*DNAcopy approach only 35 genes are found significant with an adjusted p-value smaller than 5 percent compared to 435 for the Extern approach. This difference is highly significant, Fisher's exact test:

```
## [1] 1.125581e-150
```

The following figure represents as histograms the p-values obtained comparing D1 (4C), D2 (2C) and D3 (3C) using deseq2 with the Intern (white), Extern*DNAcopy (red) and Extern (blue) normalization approaches.

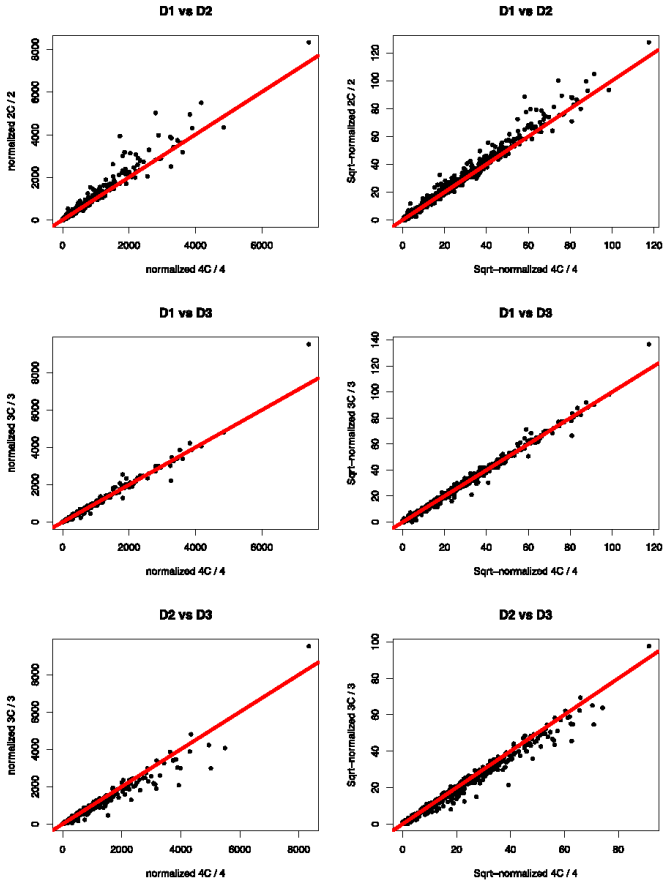


Figure 4: Normalized C count in the newly formed HE. Data have been normalized using the deseq2 size factors (obtained on genes outside of the new HE) multiplied by the C copy number (4 for sample D1, 2 for sample D2 and 3 for sample D3). The red-line represent y=x. (Left) Raw normalized count. (Right) Square rooted normalized count.

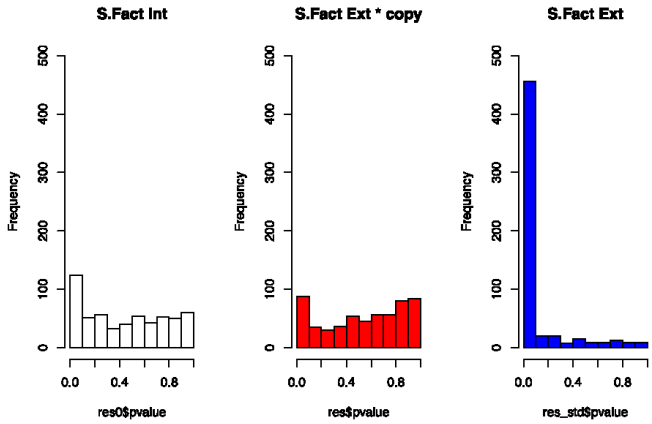


Figure 5: Histograms of p-values obtained using deseq2 on the C homeologs of the new HE genes with either the Intern (white), Extern*DNAcopy (red) and Extern (blue) size factors.

Total A+C after translocation

If we have a simple proportional DNA copy number to gene expression model any small gene expression difference between A and C prior to the translocation should lead to a change in A+C expression after the HE. For example, if A is overexpressed compared to C prior to an A translocation (2A+2C to 3A+C) we expect higher total expression of A+C count in the 3A+C sample compared to the 2A+2C sample.

To test this idea we compared the results of two deseq2 analysis: 1. (D2_AvsC) a deseq2 analysis comparing A to C in sample D2. D2 has 2 copies of A and C. 2. (D_ApC) a deseq2 analysis testing whether the total A+C of D1, D2 and D3 were equal. For both analysis we used a log-likelihood ratio test.

We then compared the list of differentially expressed genes we obtained with (D2_AvsC) and (D_ApC) using an adjusted p-value cut-off of 0.05 and 0.01. We obtained the following contingency tables and Fisher exact test p-values.

```
##           AplusC
## AvsC_in_D2 FALSE TRUE
##      FALSE   222   41
##      TRUE    148  159
```

```
## [1] 3.105242e-20
```

```
##           AplusC
## AvsC_in_D2 FALSE TRUE
##      FALSE   272   30
##      TRUE    144  124
```

```
## [1] 3.079788e-23
```

Analysis of old HE

Identification and characterization

We recovered all old HEs occurring in Yudal and discarded the "1YAn1-/1YCn1+" event. In the end we only considered a gain of A in Yudal as there very few C gains left (16 C+ vs. 472 A+).

Again we estimated the size factor using deseq2 on the none HE genes.

Analyzing A in old HE

To assess whether in the old HE gene expression is proportional to DNA copy we consider three Deseq2 analysis. In each analysis we compare A gene expression between Darmor samples having 2 copies of A and Yudal samples having 4 copies of A. We used either the Intern, Extern or Extern*DNAcopy approach.

The following figure compares for old HE genes the averaged A homeolog expression of Darmor samples and Yudal samples. The data was normalized using the Extern*DNAcopy approach. It can be seen that data-points are roughly on the $y = x$ line.

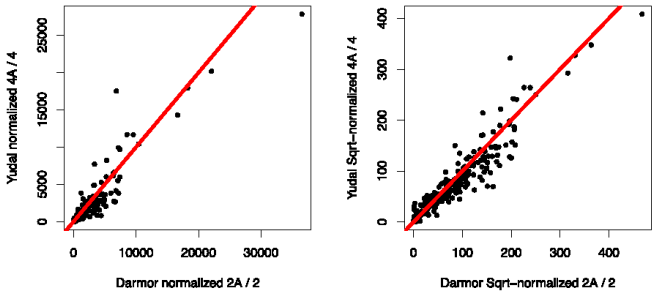


Figure 6: Normalized A count of old HE in Darmor compared to Yudal. Data have been normalized using the deseq2 size factors (obtained on genes outside of the new and old HE) multiplied by the A copy number (2 for sample Darmor and 4 for Yudal). The red-line represent y=x. (Left) Raw normalized count. (Right) Square rooted normalized count.

Out of 264 A homeologs in the old HE the p-value obtained using the Extern*DNAcopy approach is smaller than the one obtained with the Extern approach for 193 genes. 166 genes are found significant with an adjusted p-value smaller than 5 percent with the Extern*DNAcopy approach compared to 229 for the Extern approach This difference is highly significant, Fisher's exact test:

```
## [1] 2.822997e-10
```

Compared to the Extern*DNAcopy approach on new HE genes we get many more significant genes (with an adjusted p-value smaller than 0.05):

1. For old HE we have 62.88 % significant genes
2. For new HE we have only 6.13 % significant genes.

This difference is significant using a Fisher's exact test:

```
##
##  Fisher's Exact Test for Count Data
##
## data:  cbind(table(write_resA$padj <= 0.05), table(res$padj <= 0.05))
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  13.86178 33.95144
## sample estimates:
## odds ratio
##   21.45773
```

The following figure represents as histograms the p-values obtained comparing the expression of A in old HE genes in Darmor and Yudal using deseq2 with the Intern (white), Extern*DNAcopy (red) and Extern (blue) normalization approach.

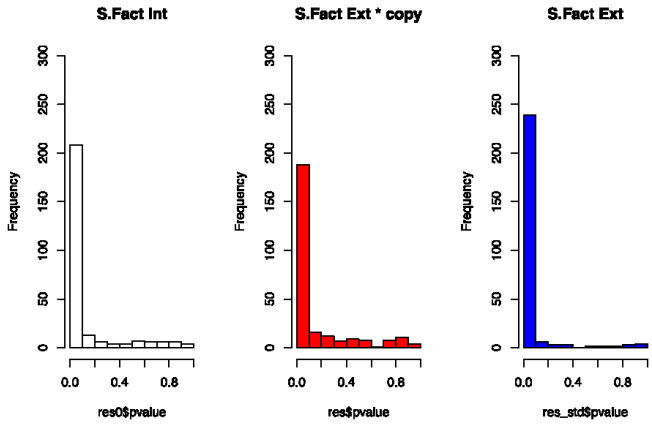


Figure 7: Histograms of p-values obtained using deseq2 to compare the A old HE of Darmor and Yudal with either the Intern (white), Extern*DNAcopy (red) and Extern (blue) size factor.

Compensation ?

We then compared the expression of A in Darmor and Yudal for old HE genes and also for none HE genes. The figure below represents the averaged expression of A (normalized with the Extern*DNAcopy approach) in Darmor and Yudal for old HE genes and none HE genes.

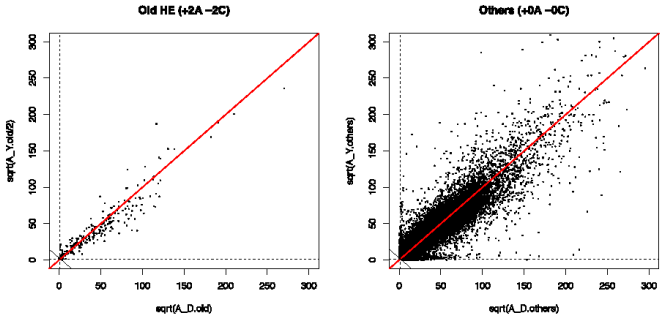


Figure 8: DNA copy number normalized A expression in Darmor compared to DNA copy number normalized A expression Yudal for old HE genes and none HE genes

Visually it seems that the distribution for old HE is less variable than for none HE genes. To assess this more precisely we looked at the angle of each point. To be specific for each gene (old HE or none HE) we computed the averaged normalized value for Yudal and the averaged normalized value for Darmor. We then took the arctangent of the Yudal-Darmor ratio and divided it by $\pi/2$ to get a value between 0 and 1. We recovered the following histograms for old and none HE genes.

A Kolmogorov-Smirnov test confirmed that the two distributions are significantly different with a p-value of:

```
## [1] 0.0003232647
```

A variance test confirmed that the variance was significantly smaller for old HE genes (0.034) compared to none HE (0.045) with a p-value of :

```
## [1] 0.002227454
```

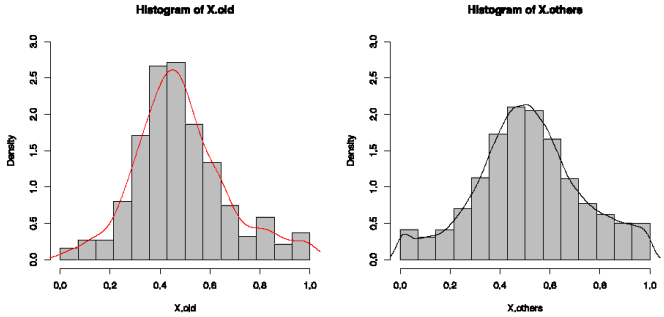


Figure 9: Arctangent of the A Yudal over A Darmor in old and none HE genes.

A and C expression in old HE and total A+C between Darmor and Yudal

We then compared the results of 2 deseq2 analysis.

1. (AvsC) We used deseq2 to compare A and C gene expression in Darmor for old HE genes.
2. (ApC) We used deseq2 to compare A+C in Darmor and Yudal for old HE genes.

Using the first analysis a gene is classified as up-regulated, down-regulated or stable we call these AvsC-up, AvsC-dw, AvsC-st. Similarly using the second analysis we classified genes in ApC-up, ApC-dw, ApC-st.

We obtained the following contingency table.

```
es1
```

```
##          i_ApC
## i_AvsC    AvsC-dw AvsC-st AvsC-up
##   ApC-dw       38       19       35
##   ApC-st       25       26       18
##   ApC-up       17       24       61
```

There are many ApC-up/AvsC-up genes compared to ApC-up/AvsC-dw genes. The number ApC-dw/AvsC-dw genes is similar to the number of

ApC-dw/AvsC-up genes. This is confirmed by a binomial test on each ApC line testing whether the proportion of AvsC-dw genes is equal or not to the AvsC-up genes. The p-values are in the table below. We also included a binomial test comparing the total number of AvsC-dw genes with the total number of AvsC_up genes. Here are the results:

```
##             p-value
## ApC-dw 7.275471e-01
## ApC-st 2.912152e-01
## ApC-up 3.588359e-07
## Sum    1.468843e-02
```

A and C expression in none HE and total A+C between Darmor and Yudal

We reproduced the same analysis on none HE genes.

We obtained the following contingency table:

```
##              i_ApC_others
## i_AvsC_others AvsC-dw AvsC-st AvsC-up
##        ApC-dw    3238    2206    3642
##        ApC-st    1630    1107    1043
##        ApC-up    3301    2142    3360
```

We compared each line of this none HE genes contingency table with the one we obtained for old HE genes using Fisher's exact test. We also compared the overall proportion of AvsC-up and AvsC-dw between old HE and none HE genes. We observed a significant p-value only for the "ApC-up" and "Sum" line.

```
### old vs others
mat_equi[1, 1]  <- fisher.test(rbind(es1[1, ], es2[1, ]))$p.value
mat_equi[2, 1]  <- fisher.test(rbind(es1[2, ], es2[2, ]))$p.value
mat_equi[3, 1]  <- fisher.test(rbind(es1[3, ], es2[3, ]))$p.value
cpt2_ <- colSums(es2)[c(1, 3)]
mat_equi[3, 1]  <- fisher.test(rbind(cpt1_, cpt2_))$p.value
mat_equi
```

```
##           p-value
## ApC-dw 0.50542287
## ApC-st 0.29607979
## ApC-up 0.01147940
## Sum    0.01468843
```

```
mat_equi
```

```
##           p-value
## ApC-dw 0.50542287
## ApC-st 0.29607979
## ApC-up 0.01147940
## Sum    0.01468843
```

C and A expression in new HE and total A+C between Yudal and Darmor

Finally we made a similar analysis for the new HE. More precisely:

1. (AvsC) We used deseq2 to compare C and A gene expression in Yudal for new HE genes.
2. (ApC) We used deseq2 to compare A+C in Darmor and Yudal for old HE genes.

We obtained the following contingency table.

```
##           i_ApC_new
## i_CvsA_new CvsA-dw CvsA-st CvsA-up
##     ApC-dw     119      71      59
##     ApC-st      39      45      49
##     ApC-up      46      56      87
```

We saw that many ApC-up genes (in Yudal compared to Darmor) are also CvsA-up (in Yudal) and that many ApC-dw genes are also CvsA-dw. This is confirmed by a binomial test on each CpA line testing whether the proportion of CvsA-dw is equal or not to the CvsA-up. Here are the p-values

```
##             p-value
## ApC-dw 6.004495e-06
## ApC-st 2.890961e-01
## ApC-up 3.594591e-04
```

Validation and denovo identification of old HE genes using gene expression data

For this analysis we kept only genes with more than 10 reads in total. We excluded the newly formed HE and ordered the genes along the genome.

Validation

We performed a deseq2 analysis to compare gene expression between Darmor and Yudal samples. For each gene we obtain a p-value and log-fold-change. Using these we computed a score per gene equal to minus the log p-value multiplied by the sign of the log-fold-change. A gene gets a high score if the gene it is significantly up-regulated in Darmor and a low score if it is significantly down-regulated in Darmor.

For each old translocation we recover nA A genes and nC C genes and computed the median score of A genes and the median score of the C genes.

We computed the median score obtained for all possible contiguous regions of nA and nC genes along the genome. We also computed the median score obtained for nA and nC randomly selected genes. We compared those scores to those obtained for the translocation to recover a p-value, more precisely:

1. for A gain in Darmor we computed the percentage of regions with a score higher than the translocation;
2. for C gain in Darmor we computed the percentage of regions with a score lower than the translocation;
3. for A gain in Yudal we computed the percentage of regions with a score lower than the translocation;
4. for C gain in Yudal we computed the percentage of regions with a score higher than the translocation.

For the 17 translocations occurring in Darmor 3 translocations were found significant using the A genes (although this is only prior to multiple testing correction when considering p-values obtained using regions), namely

1. 1DAn1+/1DCn1
2. 4DAn1+/4DCn1-
3. 6DAn1+/6DCn1-

For the 10 translocations occurring in Yudal all (except 1 for which we have no genes following filtering) were found significants (before and after Benjamini-Hochberg multiple testing correction) using the A or C genes.

Denovo Detection

For the denovo identification of old translocation here is how we proceed:

1. We ordered the gene along the genome.
2. We computed the averaged Darmor gene expression and averaged Yudal gene expression and put them to the power 1/4.
3. We compared those per chromosome using the segmentation approach described in Rigaill et al 2012
4. We obtained segments with unexpectedly high or low expression ratio. We considered segments to be part of a translocation if they had a ratio between Darmor and Yudal higher than 3/2 or lower than 2/3 and if they were made of at least 9 genes.
5. We then manually merged successive segments with high or low ratio discarding small outlier (1 or 2 genes) segments.

Biblio

(1) http://www.ncbi.nlm.nih.gov/pubmed/22796958

(2) http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html

4.3 Overview of the transcriptome of *Brassica napus* meiocytes

4.3.1 Objectives

The objective of this study is to characterise the extent to which the meiotic transcriptome of *B. napus* is variable and to identify the main sources responsible for this variation. For this we used RNA-Seq to compare the meiotic transcriptome between the two genotypes that were used to map *PrBn* (*B. napus* cv *Darmor-bzh* and cv *Yudal*; Liu et al., 2006) and that showed different progression of homologous recombination in euploids (Grandont et al., 2014). I will notably assess whether intraspecific variation of meiotic transcriptome correlates with variation in CO frequencies and whether more insight can be gained into the nature of the causal polymorphism for *PrBn*.

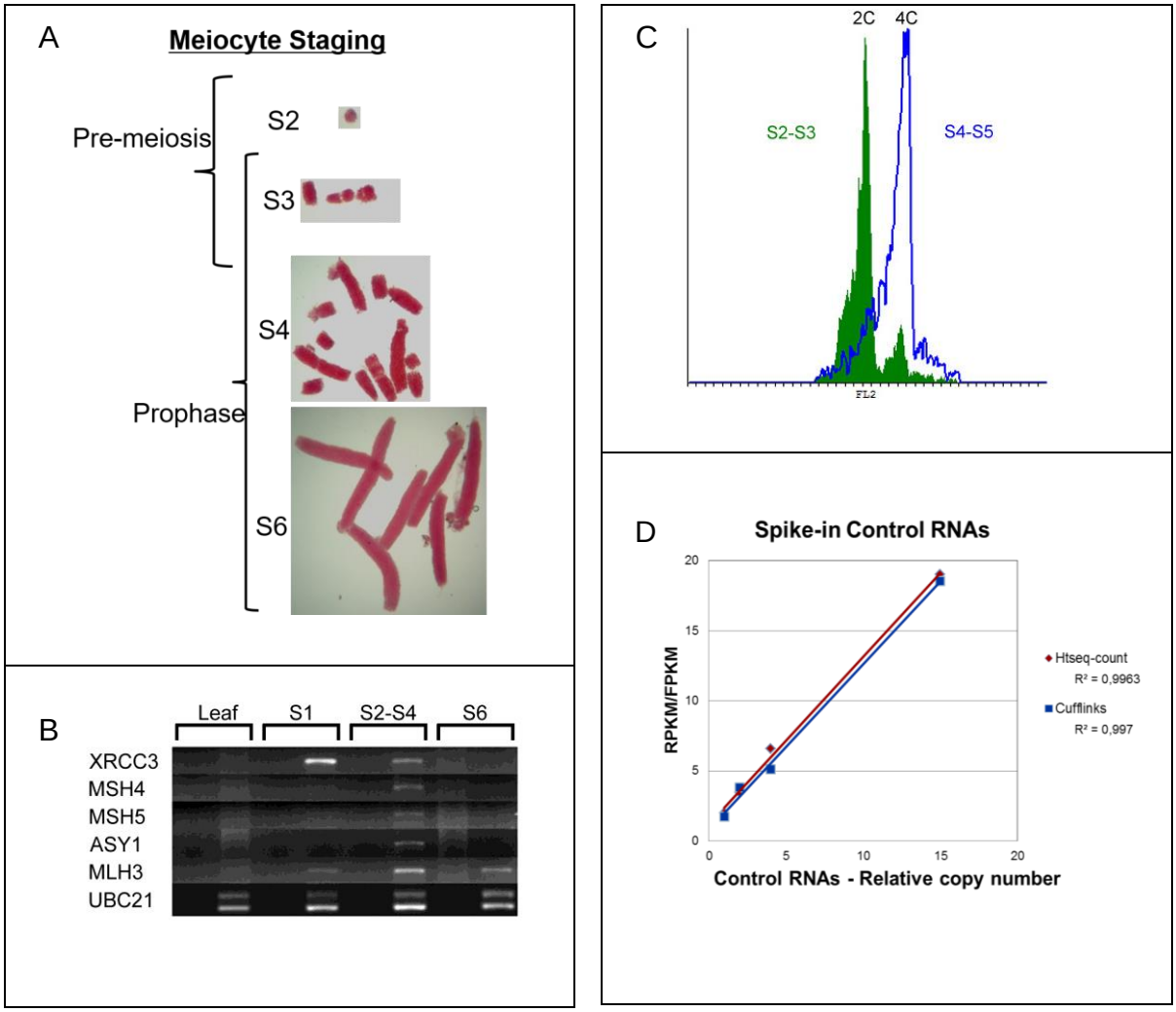


Figure 20: Setting up the experimental design

Figure 20: Setting up the experimental design

(A) Clusters of male meiocytes extracted from *B. napus* anthers. Meiocytes remain associated with one another from premeiotic S phase to the tetrad stage. The size of meiocyte clusters was used to define a series of stages (S1 to S6) from which expression of a few meiotic genes was assessed.

(B) RT-PCR assessment of total (i.e. A+C) gene expression for a subset of meiotic genes across the different stages defined above (S1, S2-S4, S6) and a non-meiotic tissue (leaf). The meiotic genes were chosen to represent different steps of meiotic recombination: chromosome axial element (*ASY1*), early DSB repair (*XRCC3*), class I CO pathway (*MSH4, MSH5, MLH3*). *UBC21* was used as an ubiquitous control for gene expression (Chen et al., 2010). For every tissue, a no reverse transcriptase control (where RT-PCR was performed in the absence of reverse transcriptase) was used to assess the amount of DNA contamination present in an RNA preparation (on the left). The products of RT-PCR performed with cDNA are shown on the right.

(C) Distribution of meiocytes (S2-S3 and S4-S5 stage are in green and blue, respectively) according to their DNA content (x-axis); y-axis represents cells counts in each classes. One main peak (left) and one secondary peak (right) are found for meiocytes in S2-S3; this indicates that only a few cells underwent DNA replication. Only the peak on the right is found for meiocytes in S4-S5 indicating that most of the cells underwent DNA replication.

(D) Correlation between transcript abundance using Htseq-counts (red) and Cufflinks (blue) (expressed in RPKM, i.e., Reads Per Kilobase of transcript per Million mapped reads or FPKM, i.e., Fragments Per Kilobase of transcript per Million mapped reads, respectively) and known molecular concentration of five control transcript in a spike-in mix.

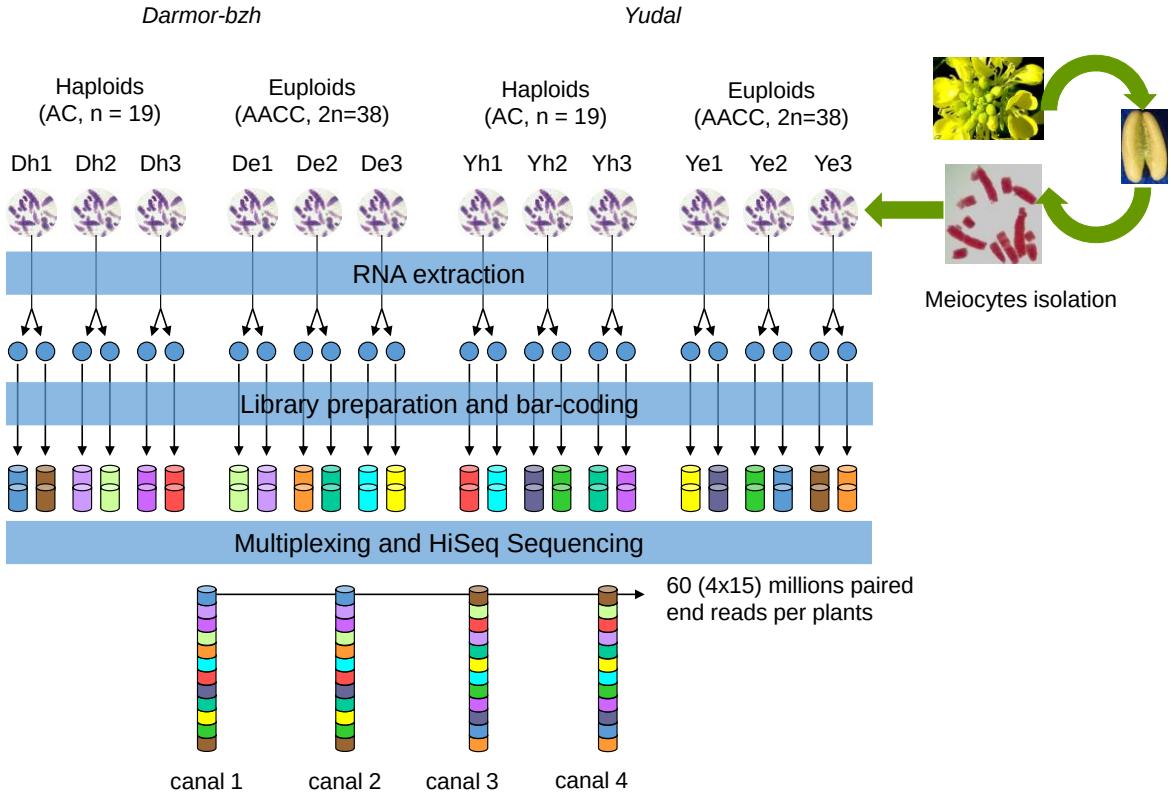


Figure 21: Experimental design for RNA-seq transcriptomic analysis

Pollen mother cells were isolated from young anthers as described in Chen et al., (2010) and used for RNA extraction. Our design includes two genotypes (*Darmor-bzh* and *Yudal*) and two level of ploidy (Euploids and Haploid), giving a total of four conditions (Dh, De, Yh and Ye). For each condition, three biological replicates (i.e. three plants) and two technical replicates per plant were used to prepare a total of 24 cDNA libraries. The two libraries from each plant were divided in two parts and sequenced on a different lane. As a consequence, each plant was represented on each lane (4 lanes in total). According to Illumina Hiseq technology specification, 60 million paired end reads (15 million per lane) are generated per plant.

4.3.2 Details on the experimental design

The reliability of our transcriptomic data depended on the correct staging of meiocytes that had to contain transcripts for a wide range of meiotic recombination genes. This was established beforehand by Andrew Lloyd using RT-PCR amplifications of genes involved in different steps of meiotic recombination. Concurrent amplification of all genes was used to define the subpopulation of cells from which RNA was extracted. Andrew then performed flow cytometry to show that meiotic gene expression occurred during DNA replication (Figure 20).

As described in Figure 21, we followed the general recommendations for proper experimental design brought up to date by Auer and Doerge (2010) for RNA-Seq experiment. Our design aimed at controlling three independent factors as well as their combined effect: (i) the sub-genomes (A or C) on which the genes are located, (ii) the genotypes (*Darmor-bzh* and *Yudal*) from which meiocytes were isolated and (iii) the level of ploidy (either allohaploids AC or euploids AACC). Thus we had a total of four conditions (i.e. *Darmor-bzh* allohaploids, DH; *Darmor-bzh* euploids, DE; *Yudal* allohaploids, YH and *Yudal* euploids, YE). For each of them, we used 3 biological replicates to estimate within-condition (biological) variability. This was indeed critical to perform differential expression analysis. Technical replicates were also included in the design (Figure 21) so much so that a sample from each plant was sequenced on each lane (4 lanes in total). By exposing equal portions of every unique sample to the same experimental conditions, balanced blocks were formed to separate the effect of the treatments from potential confounding factors (technical variations). These precautions proved to be useful in the end (see below).

Spike-in control RNAs were also incorporated in our RNA-seq experiment to measure its sensitivity and accuracy. We observed a strong correlation between read counts and RNA input over the entire detection range (Figure 20), which demonstrated that our experiment provided precise quantitative estimates of transcripts.

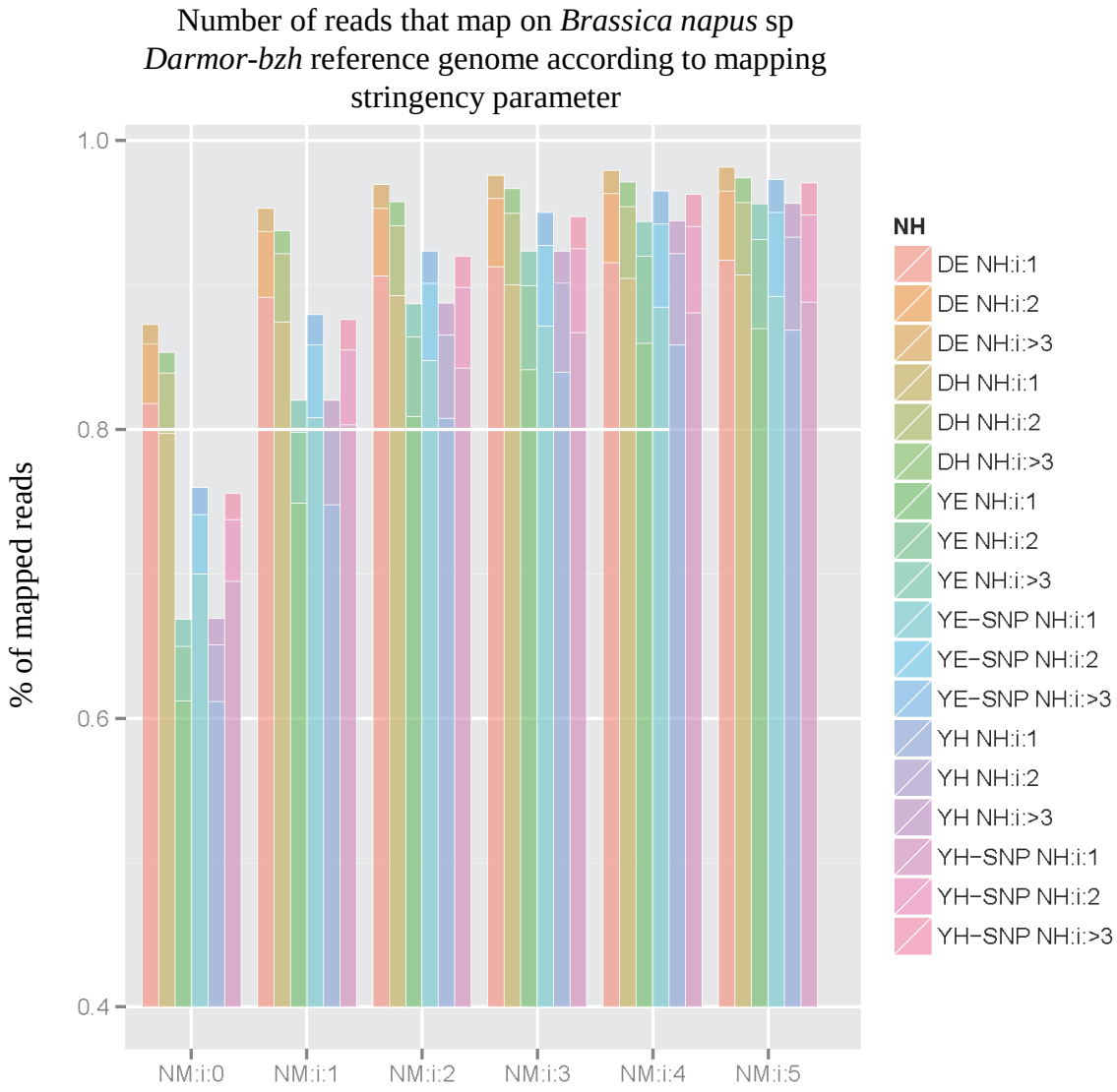


Figure 22: Cumulative proportion of reads that map on the reference genome when varying mapping stringency parameter

The proportion of reads that map for each condition (i.e., *Darmor-bzh* allohaploids, DH; *Darmor-bzh* euploids, DE; *Yudal* allohaploids, YH and *Yudal* euploids, YE) is given in % (y axis). The alignment files were filtered to assess the impact of the mapping parameters (number of mismatches i.e., NM from 0 to 5 and number of hit i.e., unique mapping (NH:i:1) or mapping at two or more positions (NH:i:2 and NH:i:>3, respectively) on mapping efficiency. For *Yudal*, SNP tolerant mapping has been performed to improve read mappability (YE-SNP and YH-SNP).

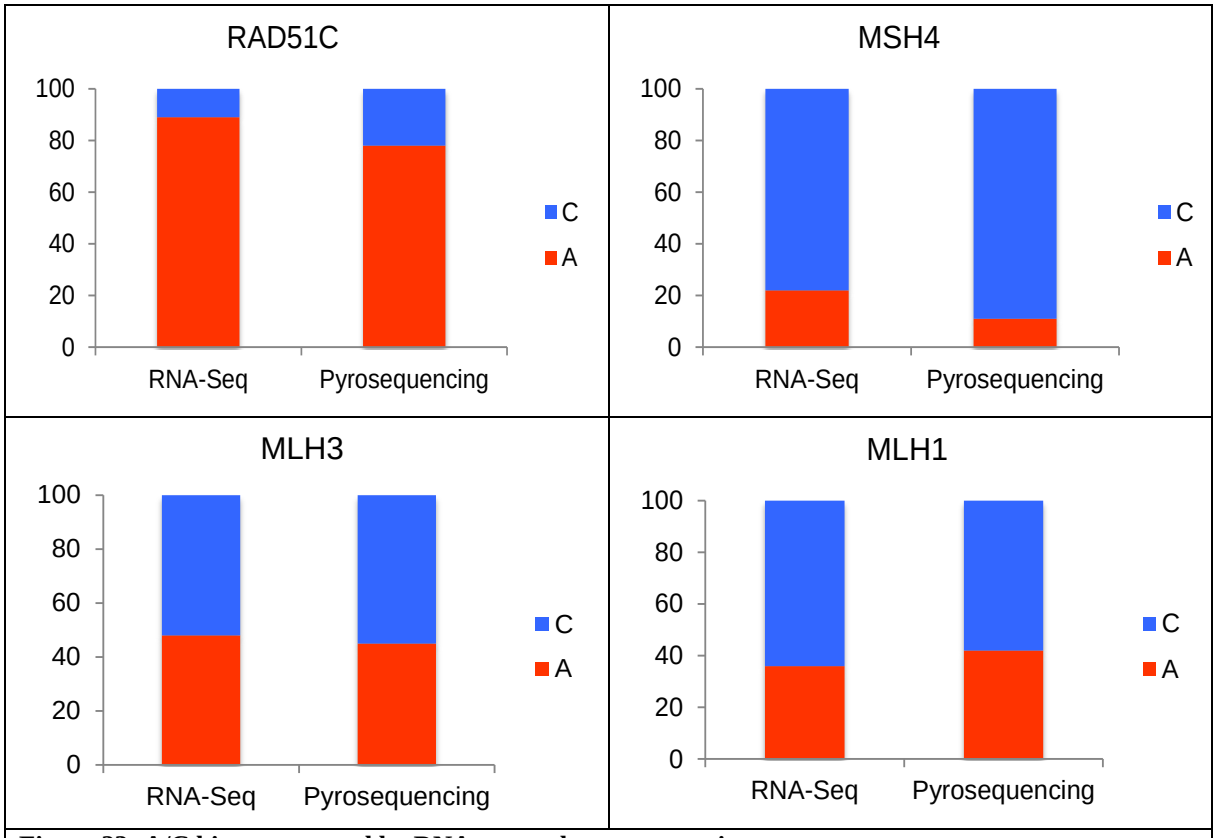


Figure 23: A/C bias as assessed by RNA-seq and pyrosequencing

Relative contribution in percent of the A (blue) or C (blue) copy to the total (A+C) expression as established by pyrosequencing or RNA-seq. RNA were extracted from carefully staged male *B. napus* meiocytes (see figure 21). Absence of preferential amplification was assessed on genomic DNA for pyrosequencing (data not shown).

4.3.3 Details on the mapping

The 24 paired-end libraries (Figure 21) were sequenced at the CNS at Evry using the Illumina Hiseq technology. After quality check and trimming (performed by the CNS), >30 million high-quality paired end reads were obtained for every sample. GSNAP was used to map reads to the *Brassica napus* cv *Darmor-bzh* reference genome sequence. To assess whether the reads coming from *Darmor-bzh* and *Yudal* map equally well on the reference genome, hereinafter referred as to mappability, we performed a first mapping at low stringency (up to 5 mismatches and multiple hits). We then filtered the resulting alignment file for the number of mismatches and for the number of hits and plotted for each condition, the cumulative percentage of mapped reads when progressively releasing these parameters (Figure 22). ~98 and 97% of reads from *Darmor-bzh* and *Yudal*, respectively mapped to the reference genome at low stringency (5 mismatches >3 hits, figure 22). By contrast, a difference of mappability was observed between *Darmor-bzh* and *Yudal* when considering uniquely mapped reads and a lower number of mismatches (Figure 22).

In order to improve the mappability of *Yudal* reads, we generated a list of SNPs differentiating *Darmor-bzh* from *Yudal* using the first mapping results (unique hit, 1 or 2 missmatch(es) for reads coming from *Darmor-bzh or Yudal* respectively). After filtering, we retained around ~292000 high quality SNPs that we used to perform a second round of mapping for *Yudal* using a SNP tolerant mapping procedure. This procedure reduced the differential mappability between *Darmor-bzh* and *Yudal* but did not completely eliminate it. In the end, ~89% and 80% of the reads from *Darmor-bzh* and *Yudal* mapped uniquely (one mismatch allowed) to the *Darmor-bzh* reference sequence, respectively. The final count matrixes were obtained using this subset of uniquely mapped reads at high stringency.

As an additional control, we used pyrosequencing on a subset of meiotic genes to verify that the reads mapped to the proper (A or C) copy of the gene they originated from (Figure 23). We also used global alignment between homoeologous gene pairs for which homoeology relationships were indisputably established by Chalhoub et al., (2014) to establish a list of ~5260000 SNP between homoeologs (homoeo-SNPs) that we used to control visually the copy specificity of the mapping for a broader sample of genes.

Table 1: Characterization of de novo transcripts via a BLAST analysis (BLAST2GO)

Non coding RNAs	
tRNAs (1)	10
miRNA (2)	10
snRNA (3) – snoRNA (4) - HACA-box	35
snRNA - snoRNA-CD-box	184
Transposable elements (Retroelements)	
SINEs (5)	11
LTR (6) elements = Gypsy/DIRS1	28

(1) transfer RNA
(2) microRNA precursors
(3) small nuclear RNA
(4) small nucleolar RNA
(5) Short Interspersed Nuclear Elements
(6) Long Terminal Repeat

4.3.4 De novo transcriptome assembly

As it currently stands, the reference genome of *B. napus* (cv. *Darmor-bzh*) contains 101,040 gene models, 90% of which have (a) clear match(es) in *B. rapa* and/or *B. oleracea* predicted proteomes (Chalhoub et al., 2014). Although a slightly larger pan-transcriptome has more recently been released for the *Brassica* A and C genomes (He et al., 2015), we used the data published by (Chalhoub et al., 2014) for our analyses, because it was the first available to us.

Among the reads that map at high stringency to the reference genome, around 7% mapped outside the existing gene models. For these reads, I ran a de novo transcriptome assembly approach (Cufflinks, Trapnell et al., 2012) to infer the corresponding transcripts by finding overlaps between the reads. This approach yielded a total of 11005 distinct transcripts with an average sequence length of 1043bp. 6426 (58%) of these newly discovered transcripts were annotated via a BLAST analysis (BLAST2GO) resulting in at least one hit to a gene or protein in the databases queried (BLAST against the NCBI non-redundant database and InterProScan tool from the EBI to retrieved domain/motif information). The remaining 4579 new transcripts were on average 688bp long. Among them, 103 transcripts did not display an open reading frame (Min et al., 2005). A BLAST against the Rfam database (using BLAST2GO interface) and prediction tools for transposable elements (LTR_FINDER, Xu and Wang (2007) allowed us to identify non-coding RNAs and transposable elements, respectively (Table 1).
These results echoed observations from *A. thaliana*, maize or sunflower meiotic transcriptomes where frequent un-annotated features were found to be transcribed in male meiocytes (Dukowic-Schulze et al., 2014a; Flórez-Zapata et al., 2014). Although expressed during meiosis, it is not known whether these genes are meiosis specific as we lack elements of comparisons in other tissues. In the current state of this study, these newly identified transcripts were not taken into accounts for further analysis.

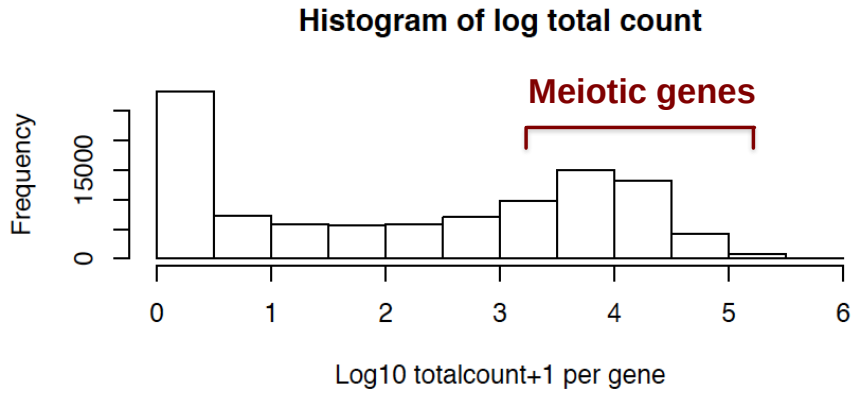


Figure 24: Gene distribution according to total read count

Histogram of the observed distribution for the total number of reads (log10) for a gene across all 24 samples. The peak around 0 corresponds to genes that are not expressed in our analysis.

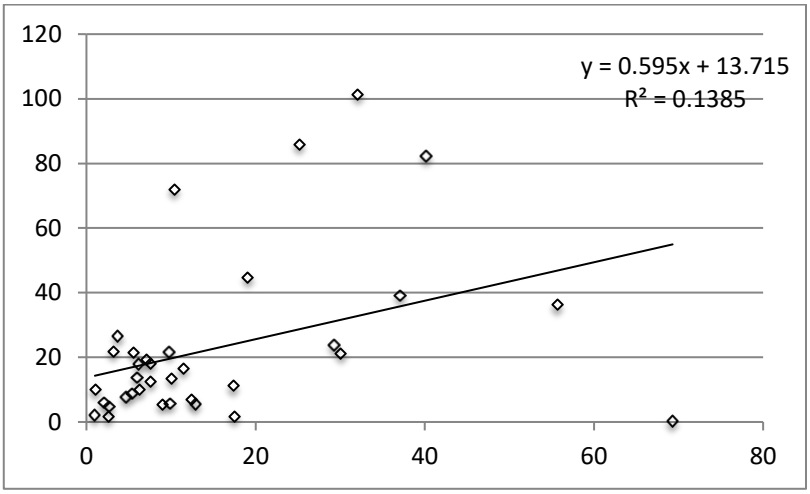


Figure 25: Correlation of meiotic gene expression in *A. thaliana* and *B. napus*

Mean expression of 35 homoeologous gene pairs in *B.napus* (y axis) plotted against gene expression of the corresponding homologs in *A.thaliana* (Chen et al., 2010). Gene expression is expressed in RPKM (i.e., Reads Per Kilobase of transcript per Million mapped reads). If a meiotic gene in *A. thaliana* had more than two homologs in *B. napus* (i.e., one homoeologous gene pair), it was not taken into account to draw this plot.

4.3.5 Description of the meiotic transcriptome of *B. napus*

We used total raw read counts to estimate the proportion of genes that were transcribed during meiosis in *B. napus*. ~19% of genes (either A or C) had a total count of reads (across all 24 samples) equal to exactly 0 and were clearly not expressed during meiosis. We found that ~68 and 45% of the analysed genes were covered by >10 and >100 reads per sample, respectively (Figure 24). This is less than observed in other tissues (96% of genes are expressed in leaves, roots, or both; Chalhoub et al., 2014)) which, unlike meiocytes, contain multiple tissue types. In line with this idea, the proportion of genes that were transcriptionally active during meiosis in *B. napus* is consistent with those measured in other plant species.

Then we looked more specifically at the expression level of meiotic genes. Based on the known meiotic genes identified in *Arabidopis thaliana* (Mercier et al., 2015), we established a list of the corresponding homologs in *B. napus* (Lloyd et al., 2014). This list contains 264 gene models in *B. napus (*122 homeologous pairs and 20 genes with no corresponding homoeologs). The vast majority of those genes were expressed during meiosis (244/260 analyzed; 94%). The distribution of total read counts for meiotic genes was very variable but clearly skewed towards high-count values (>1000 reads, Chi-2 test associated pvalue=6.5235E-23). (Figure 24). When compared to the *A. thaliana* meiotic transcriptome (established using a comparable approach; Chen et al., 2010; Yang et al., 2011), we observed that expression of meiotic genes in *B. napus* was on average twice as high as that of their homologs in *A. thaliana*, the correlation between meiotic genes expression in the two species being otherwise low (Figure 25).

Thus we were quite confident that we managed to enrich the fraction of meiotic genes in our dataset thanks to single cell-type isolation; this prompted us to examine further the extent to which meiotic transcriptome can vary between our different conditions.

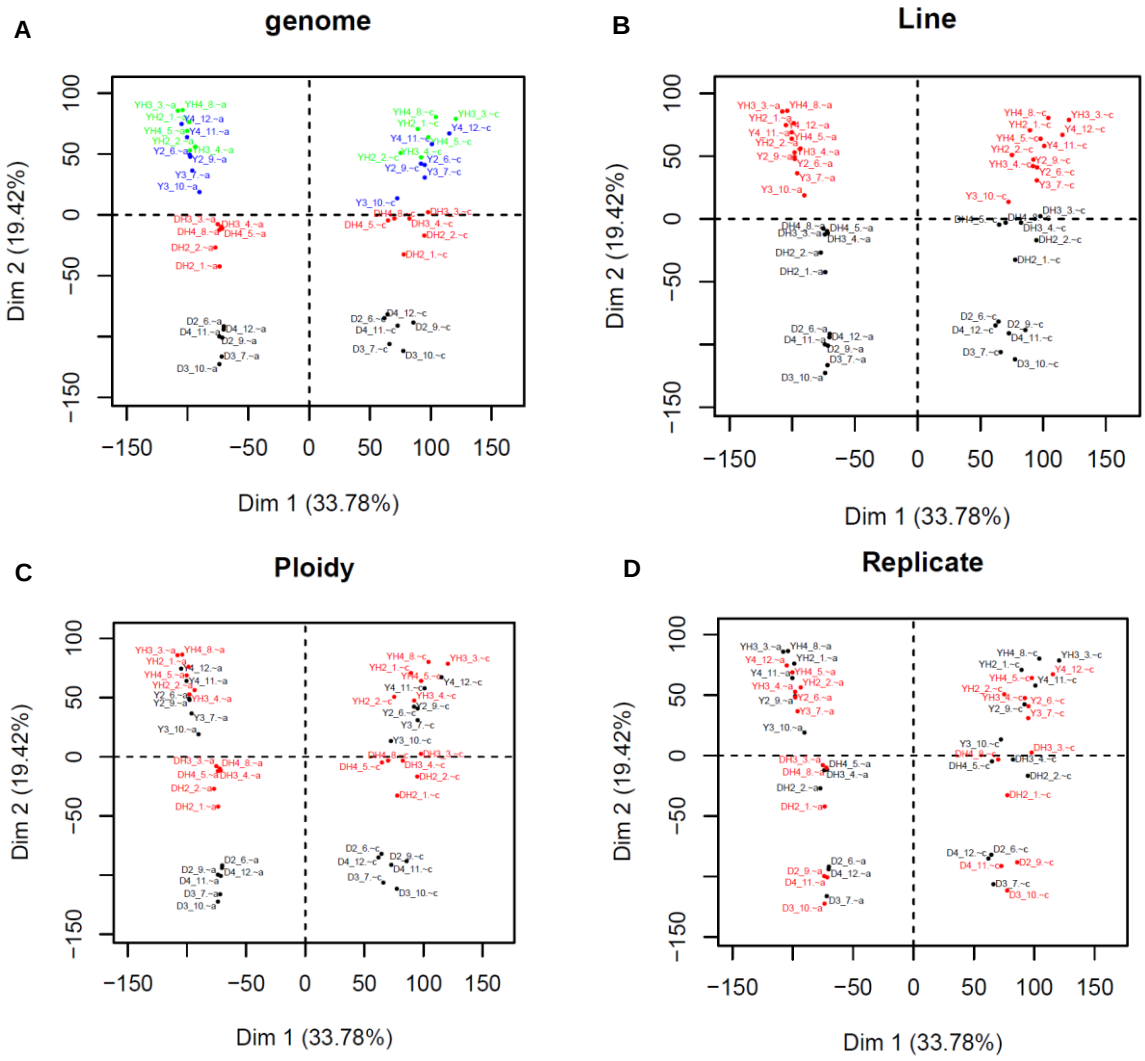


Figure 26: Score plot (homoeologous gene pairs)

The score plot displays each sample in the data set with respect to the first and second (horizontal and vertical, respectively) principal components axis. Contribution in % of each factor to the difference in variance is given for each axis (**A)** The 6 samples within each conditions are colored as follow: green = *Yudal* allohaploids, blue = *Yudal* euploids, red = *Darmor-bzh* allohaploids, black = *Darmor-bzh* euploid. (**B)** Samples are differentially colored according to the factor variety, red = *Yudal*, black = *Darmor-bzh*. (**C)** Samples are differentially colored according to the factor ploidy, red = allohaploids, black = euploid. (**D)** Samples are differentially colored according to the factor technical replicate.

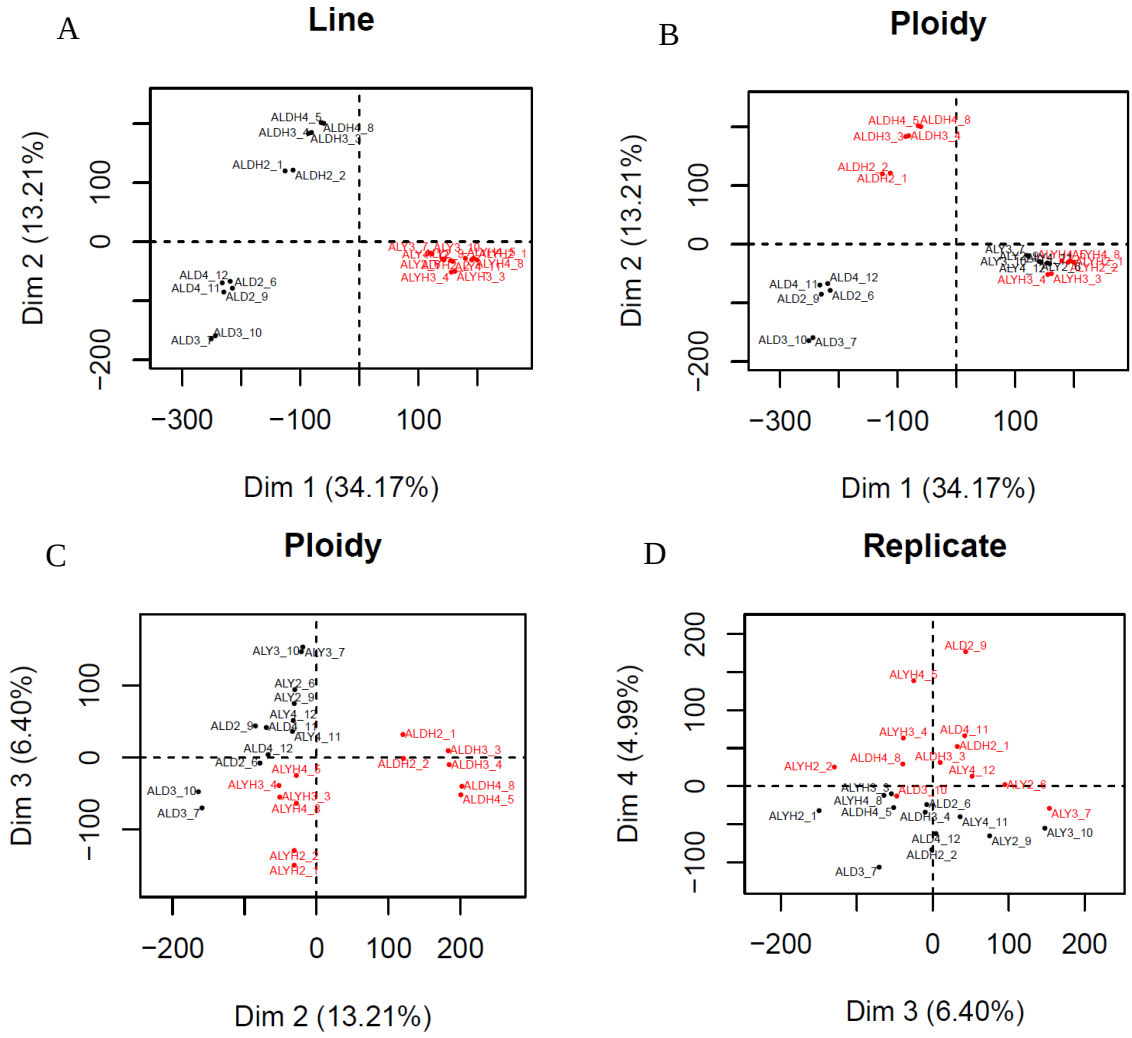


Figure 27: Score plot (entire gene set)

The score plot displays each sample in the data set with respect to the first and second (horizontal and vertical, respectively) principal components axis. Contribution in % of each factor to the difference in variance is given for each axis (**A)** Samples are differentially colored according to the factor variety, red = *Yudal*, black = *Darmor-bzh.* (**B)** Samples are differentially colored according to the factor ploidy, red = allohaploids, black = euploid. (**C)** Samples are differentially colored according to the factor ploidy, red = allohaploids, black = euploid. (**D)** Samples are differentially colored according to the factor technical replicate.

4.3.6 Partitioning the source of transcriptome variation

Next, we assessed the relative contribution of every controlled source of variation (i.e., genome, genotype, level of ploidy) to the expression of individual (A or C) genes, while taking into account possible variation between technical and/or biological replicates.

For this, we first considered expression of homoeologous gene pairs (31526 pairs for a total of 63052 out of the 101040 gene models); i.e., for which homoeology relationships were indisputably established by Chalhoub et al., (2014). In absolute terms, this conservative list remains incomplete. It is impeded both by the gaps that persist in the *B. napus* genome assembly and the strong requirements used by Chalhoub et al., (2014) to identify homoeologs (e.g., local duplicates in one genome lead to eliminate the corresponding pair from the list, as the 1:1 relationship is not satisfied). It however provided the only way to account for all sources of variation at once.

Principal components analysis (PCA) was used to explore the internal structure of our complex data set. This structure can be summarized with three principal components that explain most of the total variance in the data. In order of magnitude, the factor that accounted for most of the variation in gene expression was the sub-genome (Dim1), followed by the genotype (Dim2) and ultimately the ploidy level (Figure 26).

We repeated this PCA analysis to further explore the data using, this time, the entire gene set (101040 gene models) (Figure 27). As expected (because the effect of the sub-genome is no longer taken into account), we confirmed that the genotype now becomes the most influential determining factor for gene expression (Dim 1). Then comes the difference explained by the ploidy level; we observed that the difference between *Darmor-bzh* allohaploids and euploids (Dim2) is not the same for *Yudal* allohaploids and *Yudal* euploids (Dim3). This analysis also revealed a systematic difference between technical replicates (Dim 4). This is typical of an undesirable batch effect which is fairly common in such analysis (Auer and Doerge, 2010).

Thus we observed clear gene expression differences between samples with the 3 factors that we aimed to control explaining most of the total variance. Our experimental design was able to partition biological variation from anticipated technical variation to avoid any confounding effect.

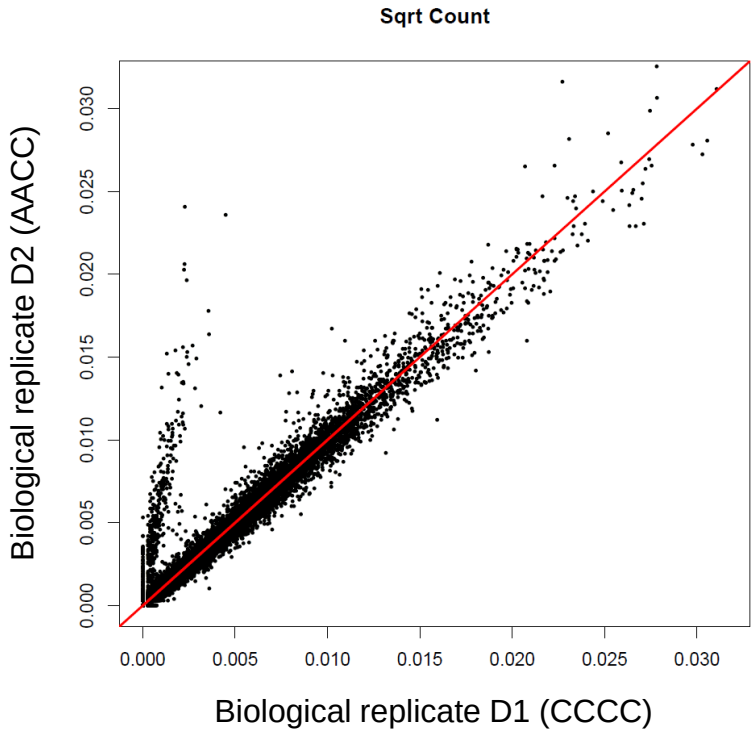


Figure 28: Detection of a newly formed HE segregating within *Darmor-bzh* euploids biological replicates
Total read count (square root) ratio between two *Darmor-bzh* euploids biological replicates (D1 and D2) across all genes from the A subgenome. D2 and D1 have lost one or two A copies, respectively in a single chromosomal region on top of A_n1-C_n1 as a result of a newly formed HE. Equivalent expression (red line) is observed between D1 and D2 except for a subset of genes within the HE that appear strongly overexpressed in D2.

All biological replicates were highly reproducible except for a region on top of A01 in *Darmor-bzh* (Figure 28). These discrepancies corresponded to the newly formed Homoeologous Exchanges (HE) that segregates among *Darmor-bzh* biological replicates (see 4.2 p.61.). The genes that fall within the border of this event (1470 genes in total) as well as within the border of all HEs that are fixed in *Darmor-bzh* and in *Yudal* (> 3,500 gene models; see 4.2 p.61.) were discarded for further analysis to avoid confusion effect.

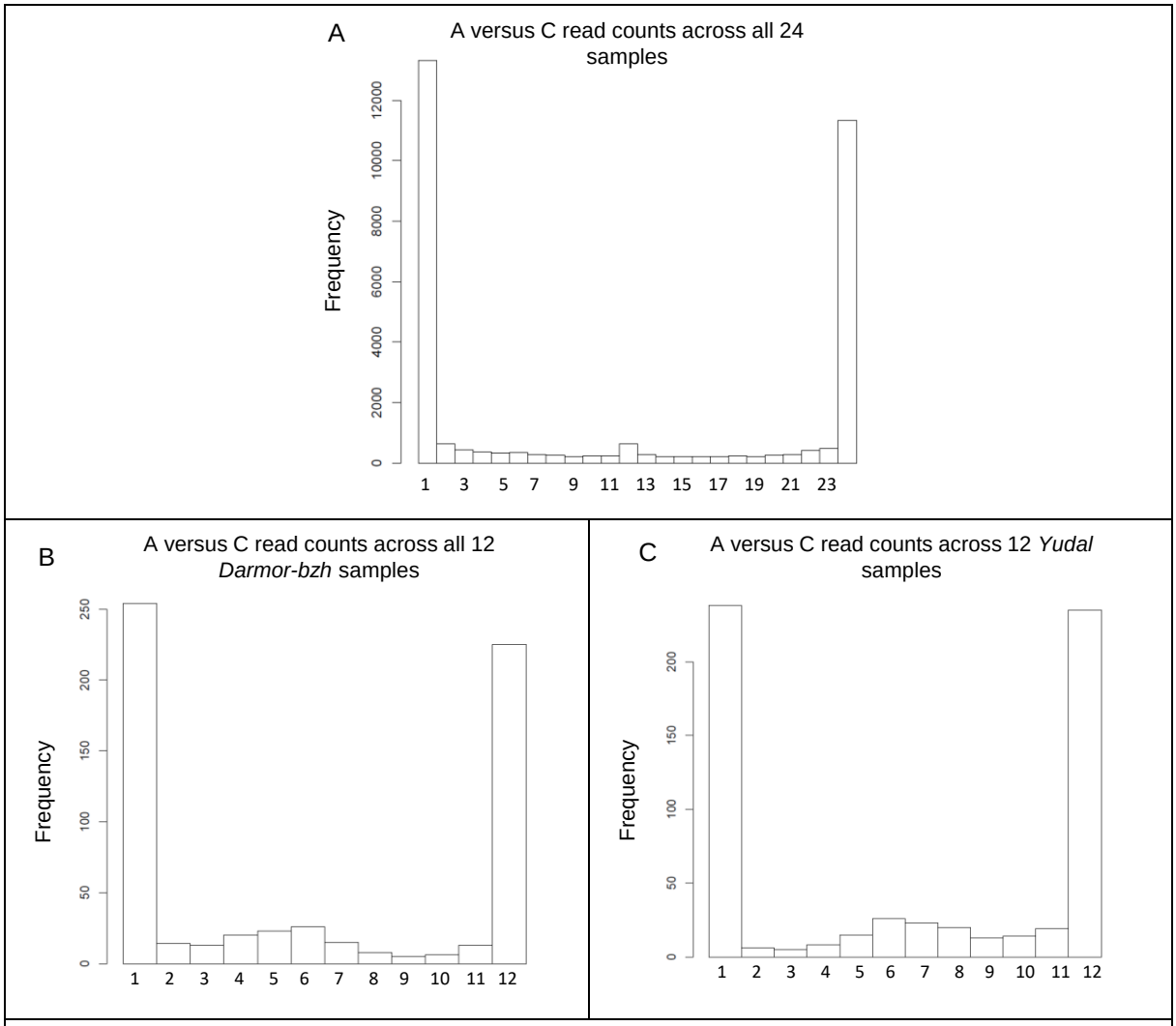


Figure 29: Consistency of total reads counts across samples

(A) 24 classes are represented along the x-axis. A given gene from a homoeologous gene pair will fall into a class depending on in how many samples total read counts for the A copy is higher than the C copy. For example, a gene will fall into class no. 1 if across all 24 samples in the experiment, total read counts for the C copy is consistently lower compared to the total read counts for the corresponding A copy. Y-axis represents how many genes fall into each class, i.e. The same analysis was repeated separately for the 12 *Darmor-bzh* samples **(B)** and for the 12 Y*udal* samples **(C)**.

4.3.7 Differential gene expression between homoeologs does not result in global sub-genome dominance

The results from the PCA analyses prompted us to examine further the sub-genome effect on gene expression. We observed that a large proportion (76%) of the genes expressed during meiosis were indeed differentially expressed between A and C copies. These variations were however of low amplitude for the vast majority (79%) of the differentially expressed homoeologous pairs (log2 (fold change) <1).

To confirm that these small variations were not just due to within-condition variations we checked the consistency of these changes across the 24 samples of the analysis (3 biological replicates and 2 technical replicates for all 4 conditions) (Figure 29).

This analysis clearly showed that differences of expression between homoeologs were consistent across all samples. Interestingly, for up to 6% of the differentially expressed homoeologous gene pairs, the direction of the change (A copy > C copy or inversely A copy < C copy) is not the same in *Darmor-bzh* compare to *Yudal*.

There was however no strong evidence that one subgenome contributed more than the other to the global expression. For 7878 pairs of homoelogs (~35%) the A copy was overexpressed over the C copy. The opposite situation, i.e., C overexpressed over A, occurred in a slightly higher proportion (~41%) (Figure 29). When looking at the subset of genes that were the most differentially expressed, we observed as many genes for which the A copy was overexpressed over the C copy (1809 gene models) than genes for which the C copy was overexpressed over the A copy (1862 gene models).

Thus, although genome is clearly the main source of variation in our dataset, there is no global dominance of one sub-genome over the other. This result confirmed and extended the observation made by (Chalhoub et al., 2014) on leaf and root transcriptomes.

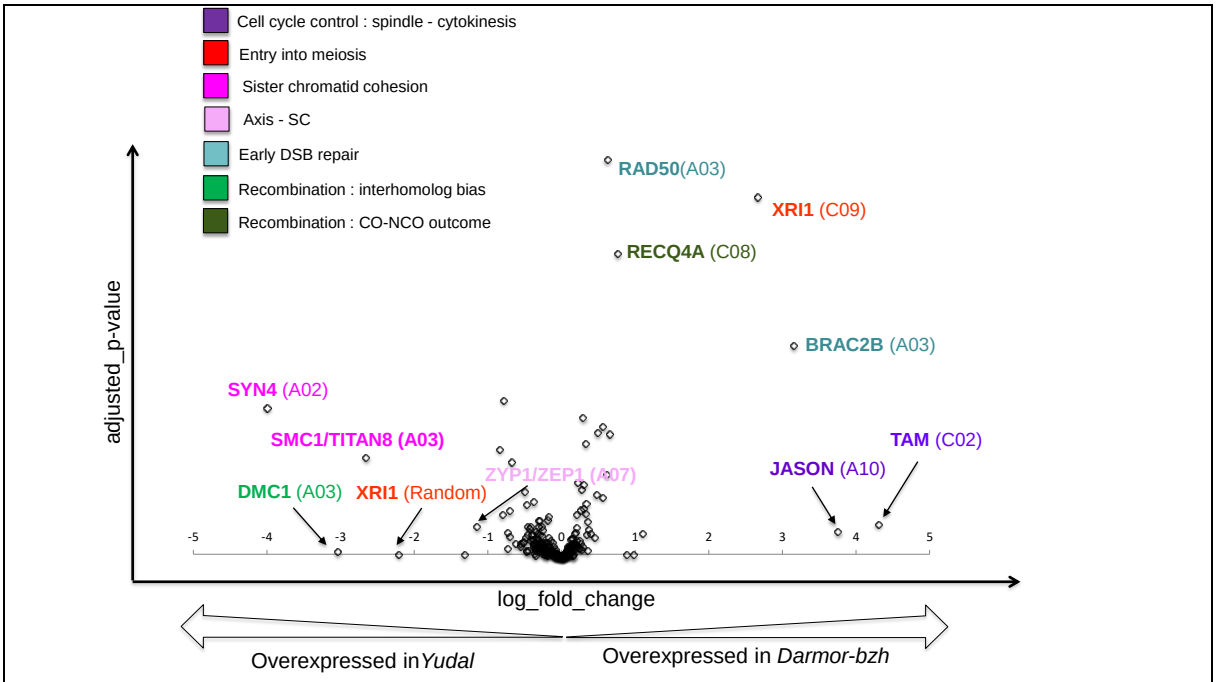


Figure 30 : Differentially expressed meiotic genes between *Darmor-bzh* and *Yudal*

Points-of-interest that display both large magnitude fold-changes (x axis) and high statistical significance (-log10 of p-value, y axis) are identified with the name of the meiotic gene (and corresponding chromosome) along with their function in meiosis (see legend). Grey arrows (bottom) indicate the direction of the change in expression (over-expressed in *Darmor-bzh* or *Yudal*, i.e., positive or negative log10 of p-value, respectively).

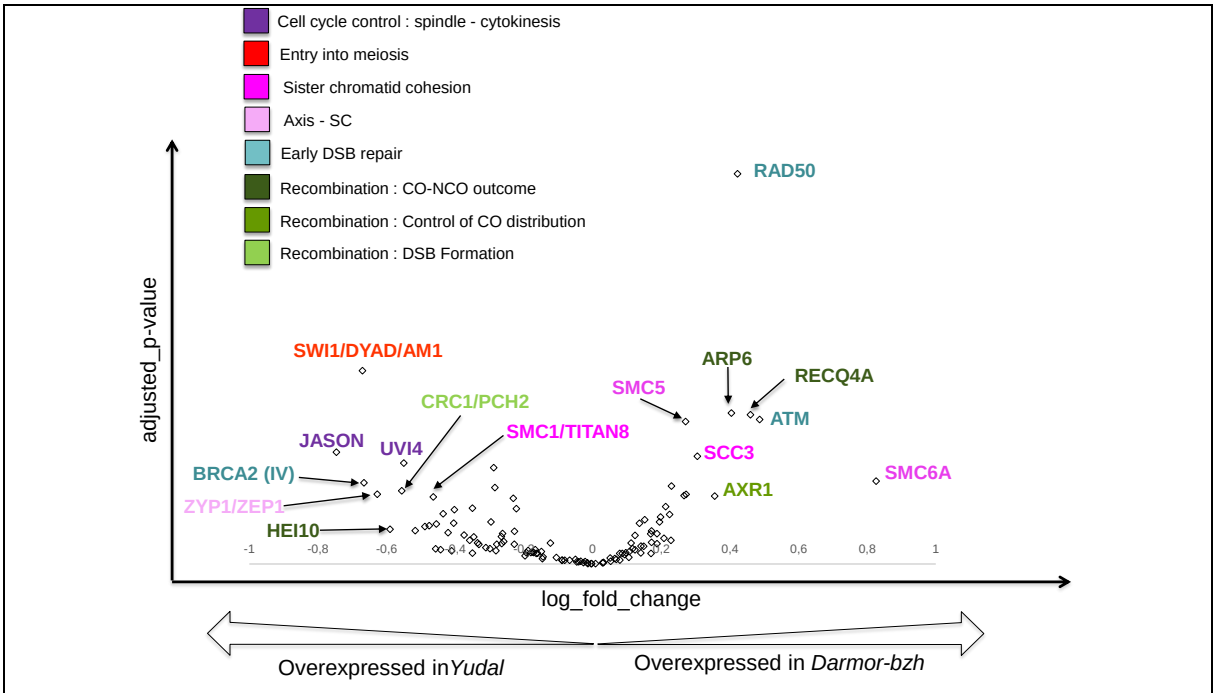


Figure 31 : Differentially expressed homoeologous gene pairs between *Darmor-bzh* and *Yudal*

Points-of-interest that display both large magnitude fold-changes (x axis) and high statistical significance (-log10 of p-value, y axis) are identified with the name of the meiotic gene along with their function in meiosis (see legend). Grey arrows (bottom) indicate the direction of the change in expression (over-expressed in *Darmor-bzh* or *Yudal*, i.e., positive or negative log10 of p-value, respectively).

4.3.8 Variation of the transcriptome between *Darmor-bzh* and *Yudal*

The second source of variation that contributed to the difference in expression was the genotype (*Darmor-bzh* or *Yudal)* (Figure 26-27). More than 60% of genes (36064 out of 60212) that were found expressed in meiocytes were actually differentially expressed between *Darmor-bzh* and *Yudal*. As per the sub-genome effect, most of these variations were of low amplitude and only 7413 genes (21%) were highly differentially expressed (log2 (fold change) >1).

This general trend held true for meiotic genes, 64% (156/242) of which were differentially expressed between *Darmor-bzh* and *Yudal* and 6% (9/156) of which showed a log2 (fold change) >1 (Figure 30). This analysis highlighted a few meiotic genes that were very significantly overexpressed in *Darmor-bzh* (*BnaA03.RAD50, BnaC08.RECQ4A* and *BnaC09.XRI1*) although with relatively low change in amplitude for *BnaA03.RAD50, BnaC08.RECQ4A.* Among the 4 genes that were found highly overexpressed in *Darmor-bzh* (log2 (fold change) >1), 3 are involved in cell cyle control (*BnaC02.TAM* and *BnaA10.JASON*) and 2 have a role in early meiosis (*BnaC09.XRI1*, *BnaA03.BRCA2*). In *Yudal*, among the 4 genes that were found highly overexpressed, 3 are involved in sister chromatid cohesion (*BnaA02.SYN4/RAD21.3* and *BnaA03.SMC1/TITAN8*) or SC formation (*BnaA07.ZYP1/ZEP1*).

We then performed the same analysis at a higher level of integration, i.e., by considering the level of expression for a given gene as the sum of the expression of its homoeologous copies (A+C). This resulted in reduced variation in expression (Figure 31); only 41% of meiotic genes were differentially expressed between *Darmor-bzh* and *Yudal*, none being highly differentially expressed. The analysis nonetheless suggested that meiotic genes tend to be more expressed in *Yudal* compared to *Darmor-bzh*. Among the most differentially expressed genes (log2 (fold change) >0.4), 76% (16/21) were overexpressed in *Yudal*, including the homoeologous gene pairs *Bna.SMC1/TITAN8* on A03/C03 and *Bna.ZYP1/ZEP1* on A07/C06. In *Darmor-bzh* we found only 5 genes that were overexpressed when summing expression of homoeologous copies, including *Bna.RECQ4A* onA09/C08 and *Bna.RAD50* on A03/C03. The difference (D>Y) observed previously for *BnaA10.JASON* was not confirmed.

It is noteworthy here that 24 meiotic genes are present in more than 2 copies in *B. napus*; this is the case *Bna.JASON* for example where the homoeologous gene pairs on A10/C05 is overexpressed in *Yudal* while the second gene pair on A08/C08 is not differentially expressed. Under the hypothesis that only the number of transcripts is relevant for the function of a gene, performing this analysis at an even higher level of integration (merging the expression data from all copies) could be informative.

Table 2: Differential expression in *B. napus* transcriptions factors that are within the most up-regulated genes in meiosis in both *Arabidopsis* and maize

	bHLH	MIKC	MYB	ERF	YABBY	bZYP	total
Darmor-bzh	0	4	1	0	1	4	10
Yudal	12	12	6	1	0	2	33

Table 3: Differential expression for *B. napus* transcriptions factors

	bHLH	MIKC	MYB	ERF	YABBY	bZYP	total
Darmor-bzh	76	27	67	37	4	57	268
Yudal	108	47	90	88	8	49	390

Altogether, these analyses showed that meiotic gene expression was variable between *Darmor-bzh* and *Yudal*, although considerably less variation was observed when considering homoeologous gene pairs. No meiotic pathway emerged as being specifically differentially expressed.

Our RNA-Seq data set could also be used to identify SNPs within meiotic genes that differentiate *Darmor-bzh* and *Yudal*. We found SNPs in *Yudal* leading to non-synonymous amino-acid changes in ~half of the meiotic genes. 2 SNPs leading to a STOP were found in the transcript sequence of *BnaC03.SMC1/TITAN8* and *BnaA03.AtSGO1*. *SMC1/TITAN8* and *AtSGO1* exist in 5 and 4 copies in *B. napus*, respectively. SNPs leading to splice variants were found in 4 genes: *BnaA05.HEi10* that acts during class I CO formation, *BnaA10.AXR1* involved in the control of CO distribution, *BnaA02.SYN4/RAD21.3* and *BnaC04.SYN4/RAD21.1* two paralogs of *REC8* involved in sister chromatid cohesion but whose exact role are not known (Zamariola et al., 2014). Except for *AXR1*, no SNP leading to non-synonymous amino-acid changes are found in the corresponding homoeologs. *AXR1* is present in 5 copies in *B. napus*, missense mutations are found in all but one copies in *Yudal*, it has not been assessed yet whether these mutations target highly conserved amino acids.

A preliminary GO term enrichment analysis was unsuccessful in detecting any enriched GO terms within the most differentially expressed genes between *Darmor-bzh* and *Yudal* (data not shown). We nevertheless examined carefully the expression of a subset of transcription factors (17 genes) which had been repeatedly found overexpressed in meiotic tissues, both in *A. thaliana* and maize (Dukowic-Schulze et al., 2014a). I first established the list of the corresponding homologs in *B. napus (*77 genes). 43 out of 77 (56%) of those genes were differentially expressed between *Darmor-bzh* and *Yudal* with a clear tendency to be overexpressed in *Yudal* (33 out of 43; 76%) (Table 2). This trend is reminiscent of the results obtained for meiotic genes. I then tested whether any *B. napus* transcription factor, whether it was preferentially transcribed in meiocytes or not, tended to be up regulated in *Yudal* compared to *Darmor-bzh*. This was the case for 55% of genes (688 out of 1244; Chi^2 P=0.007) (Table 3), a proportion that was well below that observed for meiotic genes and "meiotic" transcription factors. It remains to be established whether and the extent to which this trend may account for the genome wide differential gene expression that we observed between *Darmor-bzh* and *Yudal.*

4.3.9 Ploidy change had limited impact on meiotic gene expression

Although 16928 (33%) genes are differentially expressed between haploids and euploids, only 143 (less than 1%) of those genes showed strong variation of expression (log2 (fold change) >1). Interestingly we observed an interaction between genotypes and ploidy, the magnitude of the change between ploidy levels being far more pronounced in *Darmor-bzh* than in *Yudal* (35 and 6% of the genes are differentially expressed between euploids and allohaploids, respectively) than in *Yudal* (6%). For ~200 genes, the direction of the change in expression (euploid > allohaploid or inversely, euploid < allohaploid) is not the same in *Darmor-bzh* compare to *Yudal*.
In term of differential expression, meiotic genes followed the same general trend as presented above. Although a fair proportion (54%) of meiotic genes were differentially expressed between allohaploids and euploids, these differences were of low amplitude.

This notwithstanding, it appears clearly that the meiotic transcriptome is highly variable between *Darmor-bzh* and *Yudal* and to a lesser extent between allohaploids and euploids plants. Given what we known about natural variation of CO frequencies in *B. napus*, we wanted to have a closer insight into differential gene expression within the confidence interval for *PrBn*.

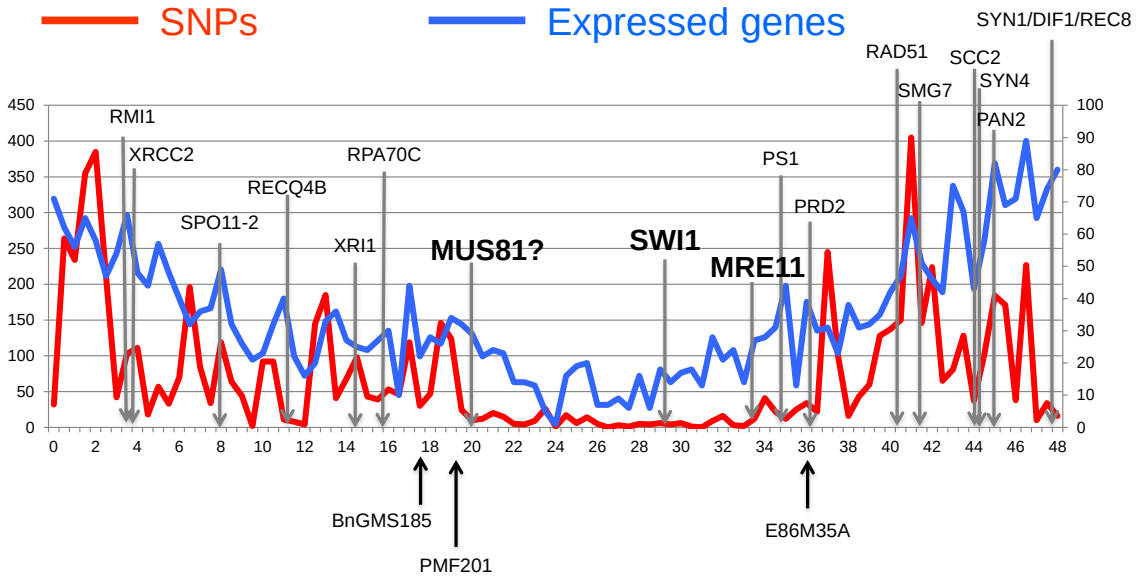


Figure 32: Genes and SNP density on the C09 chromosome

The number of expressed gene (blue – right y-axis) and the number of SNPs (red curve – left y-axis) are given for sliding windows of 500Kb along the C09 *B. napus* chromosome (x-axis). The grey arrows on top give the approximate position of the meiotic genes on C09. The black arrows on bottom give the physical position of the genetic markers most closely linked to *PrBn*. BnGMS185 and PMF201 surround the peak of the QTL.

4.3.10 A closer look into *PrBn* confidence interval

Given the low mapping resolution of *PrBn*, the markers surrounding the peak of the QTL are still far apart (>1.2 Mb) from one another (Figure 32).

The first positional gene candidate we have investigated for *PrBn* is *BnaC09-MUS81*, which is "close" to the peak of the QTL. As described in chapter 1, p12, MUS81 is essential to class II crossovers and could therefore be suspected to play a role in CO variation. The sequence of *BnaC09.MUS81* is truncated in the current genome assembly and was reconstructed before I joined the lab using BAC clones. At the very beginning of my PhD work, I used both pyrosequencing and RT-QPCR to assess (i) the relative expression of each copy and ii) the total expression of MUS81 (the primers I used were not copy-specific). I observed a very modest overexpression of *MUS81* in *Darmor-bzh*. In addition, I showed a complete absence of polymorphism in the promoter region (i.e. 10Kb upstream) of *BnaC09.MUS81* between *Darmor-bzh* and *Yudal*, extending previous observations from the gene sequence (exons+introns). I used the list of SNPs I identified between *Darmor-bzh* and *Yudal* (to improve Yudal read mapping) to test whether this absence of polymorphism is specific to *BnaC09.MUS81* or extends to a broader region. I observed that a large region of ~12Mb extending on both sides on the centromere of C09 is almost completely devoid of polymorphism, suggesting that this region had undergone a recent/strong selective sweep (Figure 32). Altogether, these results indicated that *BnaC09.MUS81* is not the causal factor for *PrBn*.

A second positional gene candidate is *BnaC09.RPA1C*, which is located within the confidence interval of *PrBn* and in a polymorphic region (Figure 32). Replication proteins A (RPA) are involved in many aspects of DNA metabolism, with RPA1C being mainly active during meiotic recombination (Li et al., 2013; Aklilu et al., 2014). The RNA-Seq data indicates that *BnaC09-RPA1C* is not differentially expressed between *Darmor-bzh* and *Yudal*, but that a non-synonymous substitution differentiates the *Darmor-bzh* and *Yudal* alleles. Now considering all genes within the 1.2 Mb long region centered on the peak of the QTL, I identified 43 genes that were differentially expressed between *Darmor-bzh* and *Yudal*, 19 of those with high difference in expression. Based on the available gene annotations, I found no obvious gene candidate for *PrBn*.

Under the hypothesis formulated in Jenczewski et al. (2003) that *PrBn* could be haplo-insufficient, I looked into this region for genes that would be both differentially expressed between allohaploid and euploid and between *Darmor-bzh* haploid and *Yudal* haploid. I found 63 expressed genes within this interval, 16 of which respected these criteria (3 with high difference in expression). Based on the available gene annotations, none of them was an obvious gene candidate for *PrBn.*

4.3.11 Conclusions and Perspectives

As this analysis is still under progress, the discussion of the results obtained so far will be presented as part of the general discussion (see paragraph 6.1 p.193). As a conclusion, I will present some analysis, that according to me, remain to be performed on this dataset.

One obvious perspective of this work is to repeat the statistical analyses with the summed expression of homoeologues (A+C) or even using an even higher level of integration; i.e. homoeologues + palealogues (which remained after the WG triplication that affected all *Brassica*). As I have already presented for the subset of genes with known function during meiosis, this procedure is expected to reduce the number of genes that are differentially expressed. However, given the overall divergence in expression between the A and C sub-genomes, I still expect to find high differences in expression between *Darmor-bzh* and *Yudal*. This leads however to the question of the level of integration that is biologically relevant. Merging expression data coming from homoeologous transcripts implicitly supposes that these transcripts are equivalent. This might well be the case for a number of genes but counter examples are bound to be found; for examples, alternative splicing has been detected in 48% of *B.napus* genes (Chalhoub et al., 2014).

Performing the differential expression analysis on the subset of non-annotated features that we found transcribed in meiocytes might also be a perspective worth exploring. Recent studies in sunflower and maize have emphasized the role of non-coding RNAs as potential regulators of meiotic gene expression (Flórez-Zapata et al., 2016; Dukowic-Schulze et al., 2016). Although these results remain highly speculative and are not well supported yet, it could be interesting to test whether the same trend are also detected in *B. napus*.

In the absence of obvious candidate genes for *PrBn*, it appears clearly that we need to improve *PrBn* mapping resolution. The production of a new genetic map using more markers (i.e. the 15K Brassica array from TraitGenetics) and more plants from the initial mapping population (Jenczewski et al., 2003) will be instrumental in that regard. On the basis of preliminary results, extensive work will be needed on the genetic map itself before performing a new QTL detection analysis.

One of our working hypothesis is that (at least some of the) distal HEs could account for part of the variation in bivalent formation between *Darmor-bzh* and *Yudal* allohaploids. Indeed, Grandont et al. (2014) observed that some homoeologous chromosome pair, like An03/Cn03 for instances, recombine more often in *Yudal* (that form 3-4 bivalents on average) than in *Darmor-bzh* (that form 6-8 bivalents on average). Interestingly we showed that an HE is fixed in *Yudal* for this chromosome pair, while no HE exist in *Darmor-bzh* (See paragraph 4.2, p.61). It is tempting to imagine that the presence of an HE in *Yudal* may act as a QTL promoting bivalent formation. Improving the resolution of the *PrBn* interval would therefore require to take these "structural" QTLs into account as covariable during the QTL analysis; however, distal HEs are usually difficult to map genetically and the current genetic map stops just downstream of the HEs. The priority is thus to find a way to map the distal HEs in order to test whether they correspond to new QTLs. This will probably require specific and/or hand-tailored mapping procedure.

Chapter 5: FANCM

In this chapter, I present the outcome of a translational biology approach to assees whether the activity of FANCM, the first anti CO protein identified in plants is conserved in *Brassica napus* and *Brassica rapa*. This chapter is presented in the form of a manuscript.

5.1 Manuscript: FANCM limits meiotic COs in *Brassica* crops

Title
FANCM limits meiotic COs in Brassica crops

Authors
Aurélien Blary[a], Adrian Gonzalo[a], Frédérique Eber[b], Laurence Cromer[a], Andrew Lloyd[a,c], Marie-Odile Lucas[b], Delphine Charif [a], Catherine Charpentier[a], Joseph Tran[a,d,e], Anne-Marie Chèvre[b] & Eric Jenczewski[a]

[a] Institut Jean-Pierre Bourgin (IJPB), INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, 78026 Versailles Cedex, France.
[b] IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, 35653 Le Rheu, France
[c] Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
[d] Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, Bâtiment 630, 91405 Orsay, France
[e] Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405, Orsay, France.

Corresponding author
Eric jenczewski: eric.jenczewski@versailles.inra.fr

Keywords

Author's contribution
The genetic and cytological assays to analyse homologous meiotic recombination have been performed in the Institut Jean-Pierre Bourgin (IJPB) by Aurélien Blary and Adrian Gonzalo, respectively. Laurence Cromer performed the targeted mutagenesis and the transformations in *A. thaliana*.
The cytological assay to analyse homoeologous meiotic recombination has been performed in IGEPP by Frédérique Eber.

Abstract

Crossing-Overs (COs) are essential for proper chromosome segregation and alleles reshuffling during meiosis. Thus, CO frequency is a limiting factor for plant breeding purposes. Over the last few years, multiple factors that limit CO frequencies have been characterized in the model plant *Arabidopsis thaliana* such as FANCM. It has not been verified whether these mechanisms are conserved in crop species.

In this study we identified EMS induced mutants for *fancm*, the first described negative regulators of CO frequencies in plants, in two species of economic relevance within the genus *Brassica*. We demonstrated that CO frequencies in the *fancm* mutant were increased both in the diploid *Brassica rapa* and in the allotetraploid *Brassica napus*. These results illustrate an example of translational biology for a trait relevant for breeding. It also brings new insights into the control of CO frequencies in an allopolyploids crop.

Introduction

Meiotic recombination is essential for proper chromosome segregation and reshuffling of genetic information through the formation of Crossing-Overs (COs); i.e., reciprocal exchanges of genetic material between homologous chromosomes. Meiotic recombination plays both a direct and an indirect role in plant genome evolution because of its inherent mutagenic nature (Rattray et al., 2015) and its influence on selection efficiency (Tiley and Burleigh, 2015). It is also central to plant breeding (Wijnker and de Jong, 2008) as it produces new combinations of alleles on which selection can act. Thus an increase in CO frequencies is predicted to result in an increase in breeding efficiency (McClosky and Tanksley, 2013). Yet the number of COs is kept low in most species, rarely exceeding 2-3 per chromosomes (Mercier et al., 2015).

COs are one of the products of meiotic recombination. Meiotic recombination is initiated by programmed double strand breaks (DSBs) (Keeney et al., 1997). DSBs are resected to form 3′ single strand DNA overhangs, which invade the intact homologous chromosome, producing D-loops that are subsequently stabilized into DNA joint molecules (JMs). Two pathways exist that convert these JMs into COs [reviewed in (Hunter, 2015)]. The first pathway, which forms the majority of COs, is dependent on a group of proteins collectively called ZMM. The distribution of the resulting “class I” CO ensures one obligate CO per pair of chromosomes and is subject to interference (the presence of one CO reducing the probability to observe another CO in the vicinity). Class I COs are marked cytologically by the MLH1 protein (Chelysheva et al., 2010).

The second pathway, which remains secondary in wild type meiosis, depends on the endonuclease MUS81; the resulting class II COs are not interferent and far more difficult to mark cytologically (Anderson et al., 2014). Yet the vast majority of DSBs are repaired as non-reciprocal exchanges of genetic material, termed non Crossing-Overs (NCOs). Because the number of DSBs vastly outnumbers COs, it has been hypothesized that negative regulators of CO frequencies exist. Purposefully designed genetic screens have thus been carried out in *Arabidopsis thaliana* and identified three distinct pathways that limit class II CO frequencies in this species (Séguéla-Arnaud et al., 2015).

The first anti-CO protein identified in plants was FANCM (Fanconi Anemia Complementation Group M) (Crismani et al., 2012). FANCM has long been recognized as a core component of the Fanconi Anemia pathway, a network of at least 17 proteins identified in human cells that preserve genome stability by promoting the processing of blocked and/or broken replication forks (Wang and Smogorzewska, 2015). In addition to a C-terminal ERCC4-like nuclease domain and a tandem helix–hairpin–helix $(HhH)_2$ domain, FANCM consists of an N-terminal bipartite SF2 helicase domain (composed of a DEXDc and a HELICc domain) (Whitby, 2010). *FANCM* orthologs have now been identified in various eukaryotes in which they do not always play the exact same role (Knoll et al., 2012).

Studies in *A. thaliana* suggest that AtFANCM has no direct role in the repair of DNA lesions but controls somatic and meiotic recombination (Knoll and Puchta, 2011; Crismani et al., 2012). During meiotic recombination, the invading strand synthetizes a small nucleotides patch using the homolog template of the intact strand. At that point, FANCM translocates along DNA and displaces the invading strand of the D-loop, allowing its annealing with the other overhang end of the DSB. This results in NCO formation through synthesis-dependent strand annealing (SDSA). Although FANCM acts as a landing pad for multiple Fanconi Anemia associated proteins, only its direct DNA-binding cofactors MHF1 and MHF2 support the FANCM anti-CO activity (Girard et al. 2014).

The SF2 helicase domain of AtFANCM appears to be critical for its anti-CO activity. Mutations in well-conserved residues of the DEXDc and a HELICc domains or in splicing sites were indeed shown to increase MUS81-dependent CO formation in *fancm* single mutants and to restore bivalent formation in *zmm* CO-defective mutants to a level indistinguishable from wild type (Crismani et al., 2012).

The boost in COs observed in *fancm* mutant, which can be up to 3.6 fold in some intervals, could be of great interest for plant breeding. Yet, to the best of our knowledge, the effect of FANCM on CO formation has never been assessed in any crop species.

In addition to the model species *A. thaliana,* the *Brassicaceae* family includes many diploid and polyploid crops (e.g., *B. rapa, B. oleracea, B. napus, B. juncea*) that show a rich diversity of morphotypes (Cheng et al., 2014). Although many of these species can be used as a vegetable, fodder, oilseed or even as ornamental crops, diploid *B. rapa* (chinese cabbage, turnip, pak choi…) and *B. oleracea* (cabbage, Brussels sprouts, broccoli, cauliflower…) are often referred to as leaf vegetables while allotetraploid *B. napus* (oilseed rape or canola) is mainly cultivated as an oilseed crop. *B. napus* (AACC; 2n=38) arose from multiple hybridization events between the ancestors of modern *B. oleracea* (CC; 2n=18) and *B. rapa* (AA; 2n=20). Because the A and C genome progenitors of *B. napus* have experienced a whole-genome triplication (WGT) before hybridisation (Lysak et al., 2005), every gene in *A. thaliana* could possibly have up to 6 homologs in *B. napus.* Such a high number of homologs is rarely observed as fractionation, the process by which additional gene copies are lost (Freeling, 2009; Woodhouse et al., 2010), starts right after the onset of WGD (Li et al., 2016). The trend is especially strong for meiotic recombination genes that return to a single copy more rapidly than genome-wide average in angiosperms (Lloyd et al., 2014).

Intense selection in *Brassica* resulted in a notable decline in genetic diversity in modern cultivars of *B.napus* (Hasan et al., 2006; Qian et al., 2014), *B. rapa* and *B. oleracea* (Cheng et al., 2016). Increasing meiotic recombination in *Brassica* crops could thus be of great interest to reintroduce allelic diversity in these cultivated species. In this study, we explore the anti-CO activity of FANCM in two *Brassica* species, the diploid *B. rapa* and the allotetraploid *B. napus,* as a proof-of-concept for all the other crops of this family (and maybe beyond).

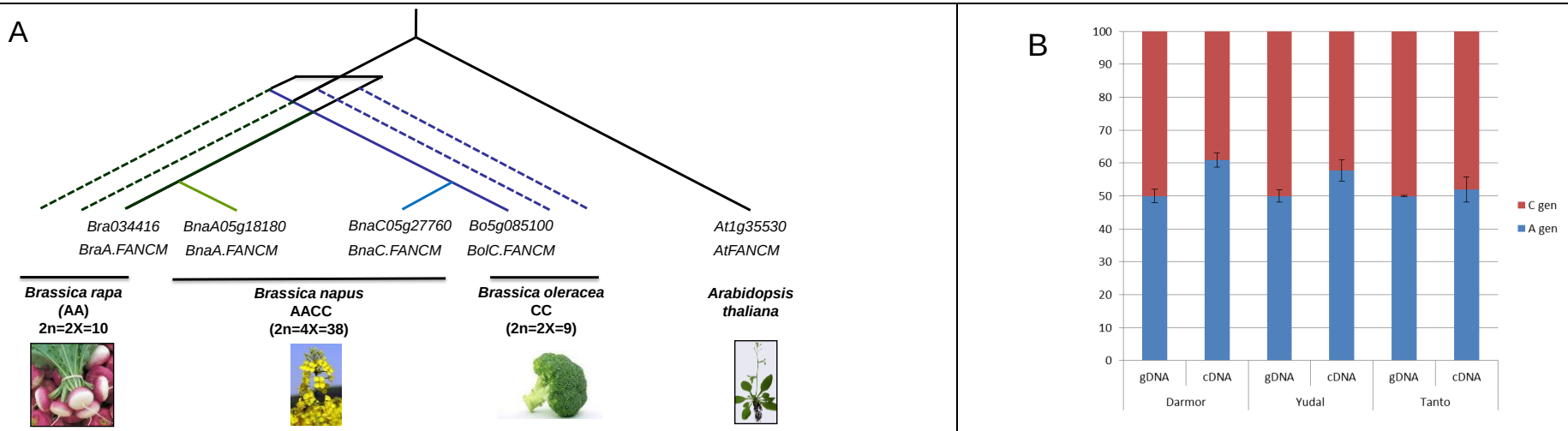


Figure 1: One copy of *FANCM* per subgenome is present and expressed in *Brassica napus*

(A) Representation of *Brassica* specific whole genome triplication (WGT) and subsequent gene loss through fractionation (dotted lines) that have experienced the homologs of FANCM in *Brassica* compared as in *Arabidopsis thaliana.* As a result of these processes, only FANCM is found in one copy in modern *B. oleracea* and *B. rapa* and in two copies (one per subgenome) in *B. napus*.

(B) Pyrosequencing gives access to the relative contribution (in percent) of *BnaA.FANCM* (blue) and *BnaC.FANCM* (red) to the total expression (A+C) of FANCM in three varieties of *B. napus*. Genomic DNA (gDNA) is used to control the absence of preferential amplification.

Results

FANCM is present in one single copy *per Brassica* sub-genomes

We first assessed the number of copies of *FANCM* that were retained in each Brassica sub-genomes after the WGT they all experienced. Querying the CDS of *At.FANCM* (JQ278026) against the available genome sequences revealed that *FANCM* has one single homologue in both *Brassica rapa* (*Bra034416* on chromosome A05, hereinafter referred as to *BraA.FANCM*) and *B. oleracea* (*Bo5g085100* on chromosome C05 = *BolC.FANCM*) while *B. napus* contains the additive gene content of its two progenitors (Figure 1A). The presence of two *FANCM* homologues in *B. napus* (*BnaA05g18180D*/*BnaA.FANCM* on A05 and *BnaC05g27760D*/*BnaC.FANCM* on C05) was further confirmed by BAC screening and sequencing. These additional sequences were instrumental to complete the full-length sequences of *BnaA.FANCM* and *BnaC.FANCM* that are still pending in the published assembly. These two genes are located within syntenic regions and therefore form a pair of homoeologues (see Table S19 in Chalhoub et al., 2014). We used mRNA-Seq data produced from *B. napus* male meiocytes (Blary, Lloyd et al., in prep) to show that *BnaA.FANCM* and *BnaC.FANCM* are almost equally transcribed during meiosis in this species; this result was subsequently confirmed by pyrosequencing (Figure 1B).

BnaA.FANCM and *BnaC.FANCM* have almost the same intron/exon structure; they only differ by the presence of a small (70bp) additional intron in *BnaC.FANCM* (and *BolC.FANCM*) that split Exon 2 in two parts. The two predicted proteins share >97% identity and >96% similarity across their full length. They are highly related to At.FANCM (~81% identity and ~84% similarity with JQ278026), in particular in the regions of the DEXDc and a HELICc helicase domains (Figure S1).

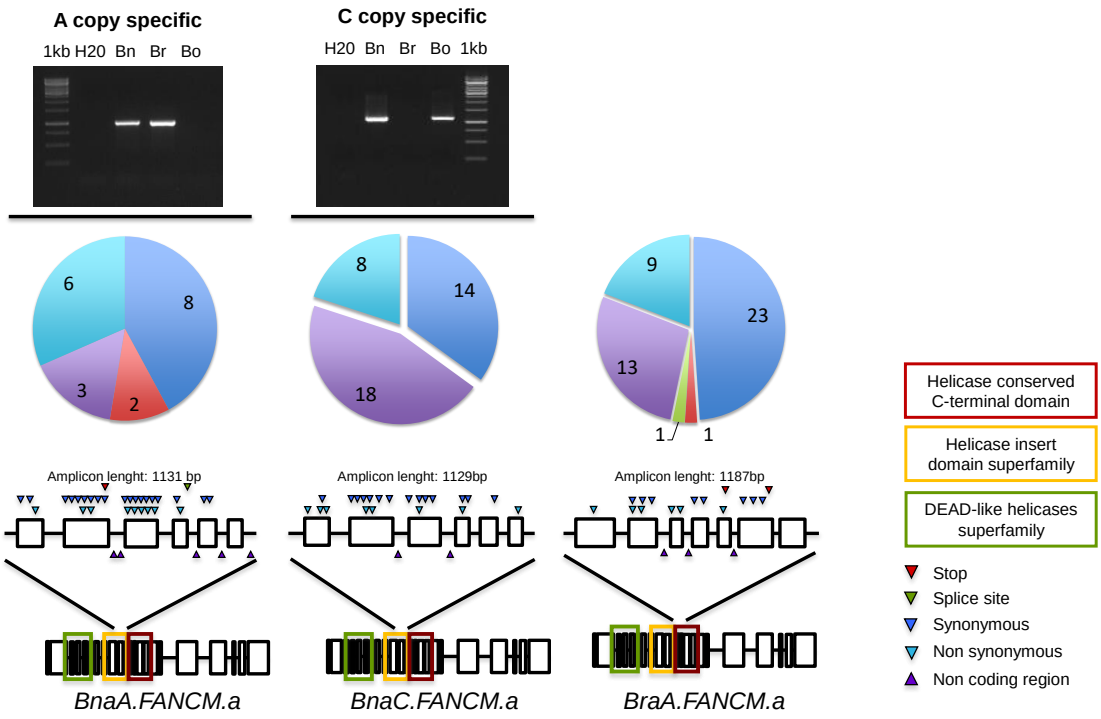


Figure 2: Outcome of three independent screens performed through TILLING to detect mutations in *FANCM* in *Brassica* species

For each homolog of *FANCM*, the composition intron-exon of the genes models is represented along with the approximate position of the helicase domains. The mutations that have been sequenced are represented along the regions (around 1kb long) that have been targeted to perform the TILLING experiment (see legend). The number of detected mutations (not necessarily sequenced) is given (pie charts). The amplification of *BnaA.FANCM* and *BnaC.FANCM* has been performed with two copy specific primer pairs that amplify either the targeted region in the A subgenome of *Brassica napus* (Bn) and the corresponding region in *B. rapa* (Br) or the targeted region in the C subgenome of *B. napus* and the corresponding region in *B. oleraceae* (Bo).

EMS mutagenesis yielded point mutations predicted to alter the function of FANCM in *Brassica*

The presence of ≤ 2 copies of *FANCM* in *Brassica* made it possible to perform TILLING (Targeting-Induced Local Lesions IN Genomes) to identify mutations in these genes. Two EMS (Ethylmethanesulfonate) mutagenized populations (one for *B. rapa* and one for *B. napus*; ~500 M2 plants each) were screened for mutations within ~1kb of the bipartite helicase domain of FANCM (Figure 2) where many loss-of-function mutations are concentrated in *A. thaliana* (Crismani et al., 2012). Two separate screens, based on the use of copy-specific primer pairs (Figure 2), were carried out in *B. napus* to find mutations affecting specifically *BnaA.FANCM* or *BnaC.FANCM*.

In total, 106 mutations were identified over all the three genes, with considerable gene-to-gene variation (Figure 2); i.e., EMS mutations were found every 12, 14 and 31 Kb in *BnaA.FANCM, BnaC.FANCM* and *Bra.FANCM* respectively. Around half of these mutations (57/106) were synonymous substitutions or occurred in introns (Figure 2).

For *BraA.FANCM*, 3 mutations within the HELICc domain were retained for further analyses, but only one was used in the present study for lack of time. The missense mutation (R54A referred latter to as *braA.fancm-1)* consists of a substitution of a strongly conserved proline at position 443 into a leucine (Figure S1). The two additional mutations, R25A and R75A induce stop codons at position 507 and 559, respectively.

For *BnaA.FANCM*, only one non-sense mutation (N84A, *bnaA.fancm-1)* was retained; it induced a premature stop codon, in-between the DEXDc and the HELICc domain. By contrast, no non-sense or splice site mutations were identified for *BnaC.FANCM*; we therefore retained two missense mutations that targeted highly conserved amino acids (Figure S1). N23C (hereinafter referred to as *bnaC.fancm-1*) and N67C (*bnaC.fancm-2*) consisted of substitutions of a leucine into a phenylalanine and a glycine into an arginine, respectively. Interestingly substitution of the same glycine into a glutamic acid was shown to be causal for a defective FANCM protein in *A. thaliana* (Crismani et al., 2012).

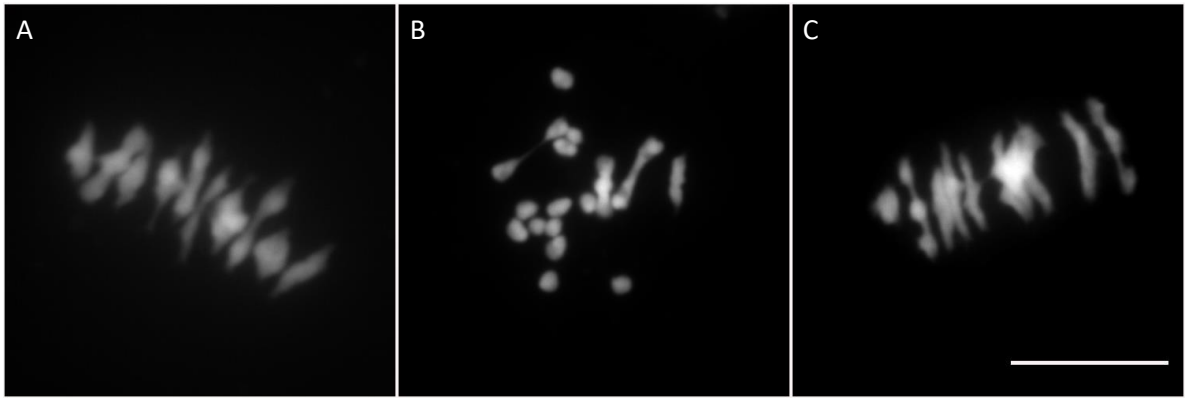


Figure 3: Restoration of bivalent formation in the double mutant *braA.msh4-1*$^{-/-}$ *braA.fancm-*$^{-/-}$

(A) During metaphase I in wild-type *B. rapa*, 10 bivalents and no univalent are formed. They are all aligned on the metaphase plate. **(B)** In the single *braA.msh4-1*$^{-/-}$ mutant, only a few bivalents are formed, most of the chromosomes remain as univalents. **(C)** Metaphase I in the double mutant *braA.msh4-1*$^{-/-}$ *braA.fancm-*$^{-/-}$ is reminiscent of metaphase I in wild-type *B. rapa*, mostly bivalents are formed, only ~0.5 univalent pair is found on average per cell. Scale bar = 10um

FANCM limits CO frequencies in *Brassica rapa*

To test whether FANCM limits COs in *B. rapa*, we replicated the cytological assay that was used to first identify the anti-CO protein activity of this protein in *A. thaliana* (Crismani et al., 2012); i.e., we tested whether *bra.fancm-1* was able to restore bivalent formation in an otherwise CO-defective mutant. As a prerequisite for this analysis, we identified through TILLING a deleterious mutation in *BraA.MSH4*, the single copy homologue of *AtMSH4* (Lloyd et al., 2014) that encodes an essential ZMM protein (Higgins et al., 2004). The mutation *braA.msh4-1* induced a substitution in the acceptor site of the 19th exon (*BraA.MSH4* has 24 exons) right after position 628 (Figure S2). We then tested whether *braA.msh4-1* was a loss-of-function mutation. In the single homozygous mutant for *braA.msh4-1* (*braA.msh4-1*$^{-/-}$) we observed numerous univalents at metaphase I, which were reminiscent of the meiotic behaviour of *Atmsh4* single mutant (Higgins et al., 2004); a mean number of only 3.7 bivalents (n=44 cells) and 4.05 ± 1.82 chiasmata (n=44), i.e., the cytological manifestation of meiotic COs, were observed in *braA.msh4-1*$^{-/-}$ compared to 10 bivalents (n=66) and 14.8 ±1.5 chiasmata (n=35) in the wild type (WT), respectively (Figure 3A-B). This demonstrated a shortage in CO formation in *braA.msh4-1*$^{-/-}$. We then produced a double mutant plant for *BraA.MSH4* and *BraA.FANCM* (*braA.msh4-1*$^{-/-}$ *braA.fancm-1*$^{-/-}$) and assessed meiotic recombination frequencies using the same cytological approaches. We observed a large increase in bivalent formation (9.44 bivalents per PMC on average; n=66) in *braA.msh4-1*$^{-/-}$ *braA.fancm-1*$^{-/-}$ and chiasmata (14 ± 2.8; n=34) in the double mutant *braA.msh4-1*$^{-/-}$ *braA.fancm-*$^{-/-}$ as compared to wild-type cells. Thus the number of bivalent and chiasmata in the double mutant *braA.msh4-1*$^{-/-}$ *braA.fancm-*$^{-/-}$ were almost indistinguishable for that of the WT (Figure 3A-C) (see above). The observation of a small univalent frequency (0.57 univalent per cell) suggests a random distribution of CO consistent with the absence of obligate class I CO (Crismani et al., 2012). Altogether, these results indicated that BraA.FANCM, like At.FANCM, limits CO formation.

Setting up a genetic assay to analyse recombination in *Brassica napus*

Replicating the experimental assay described above for *B. rapa* is hardly feasible in *B. napus*, in which it would require combining (at the homozygous state) mutations for four genes (2 copies of *FANCM* and 2 copies of *MSH4*; see Lloyd et al., 2014), instead of two. Rather, we used a genetic assay to assess the effect of FANCM on CO frequencies in *B. napus*.

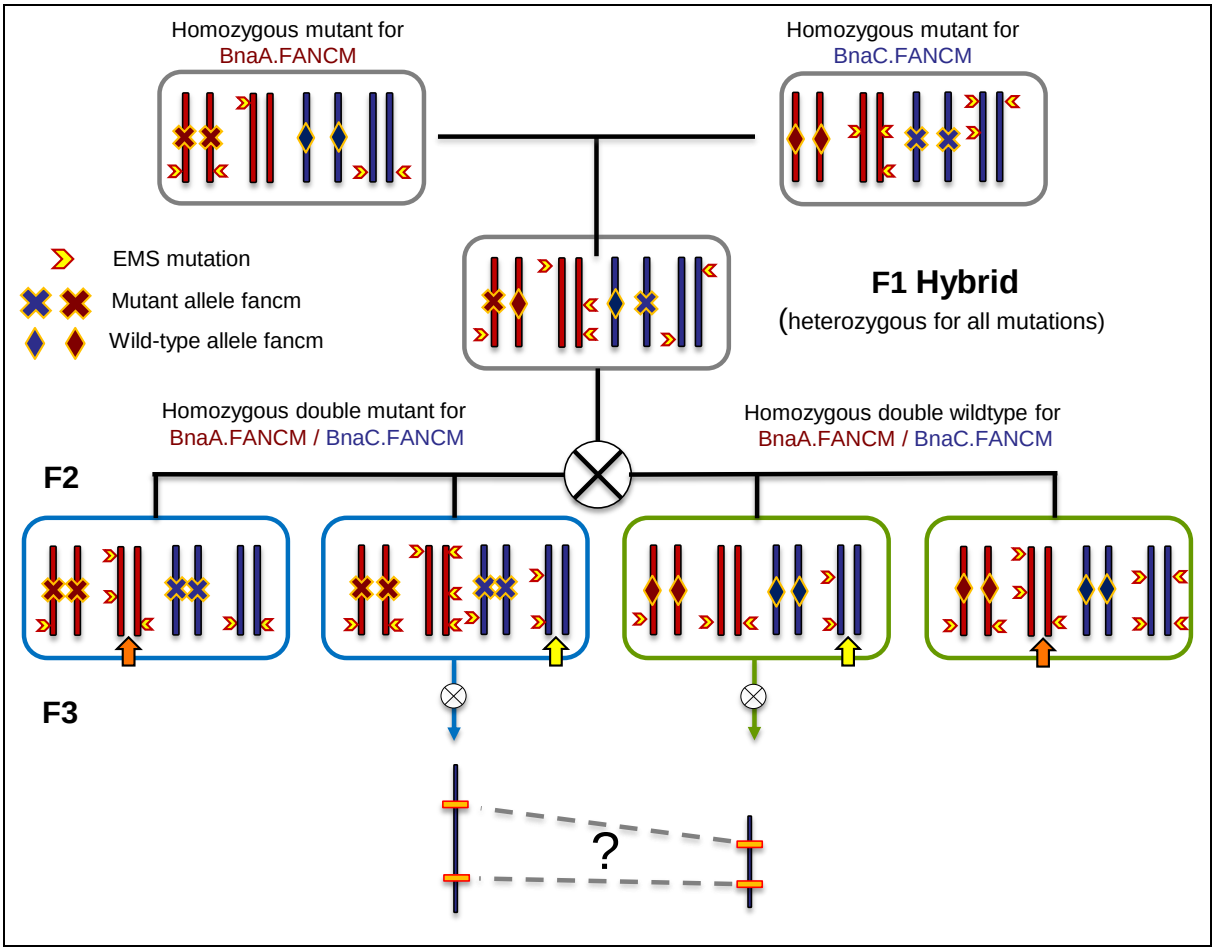


Figure 4: Experimental design to assess homologous recombination frequencies in *Brassica napus*

For each plant, the allelic version for *BnaA.FANCM* and *BnaC.FANCM* on chromosome A05 and C05 respectively, are shown as well as one random additional homologous chromosomes pairs for each A (red) and C (blue) subgenome (the sister chromatids are not represented). Each parental plants, either single homozygous mutant for *BnaA.FANCM* or for *BnaC.FANCM*, displays a unique set of EMS mutations (EMS-SNPs). After crossing, the EMS-SNPs segregates in the F1 hybrid which is heterozygous for all mutations. Each parental plant is sequenced to detect the phase (coupling-repulsion) of the EMS-SNP. Two F2 plants homozygous double mutant for *FANCM (*blue square*)* and two F2 plants double wild-type for *FANCM* (green square*)* are sequenced as well to detect common EMS-SNPs at the heterozygous stages (orange and yellow arrows). These common EMS-SNPs are used to define intervals in which recombination frequencies are compared in the F3 progenies (genetic map).

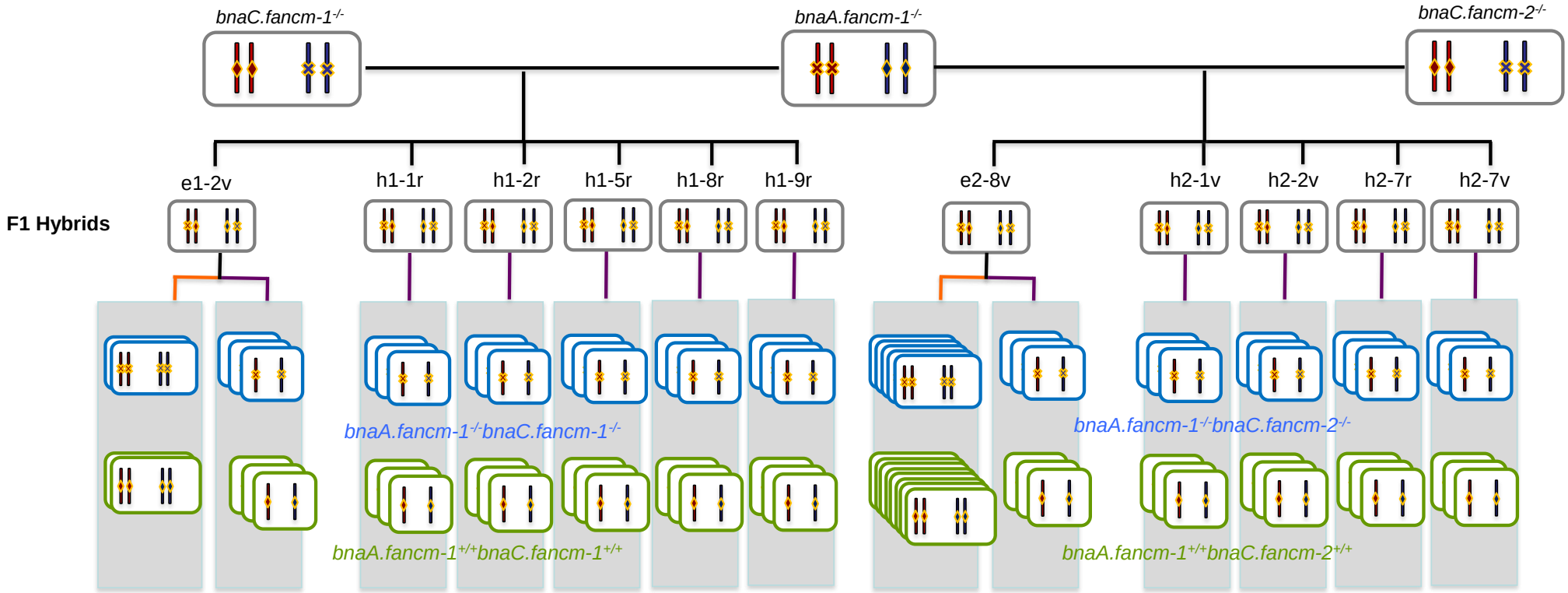


Figure 5: Genealogy of the lines used in the study

For each plant, the allelic version for *BnaA.FANCM* and *BnaC.FANCM* on chromosome A05 (red) and C05 (blue) is shown (the sister chromatids are not represented).

Multiple double homozygous mutant for *FANCM* (*bnaA.fancm-1*$^{-/-}$*bnaC.fancm-1*$^{-/-}$ or *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$*)* (blue square*)* and double wild-type for *FANCM (bnaA.fancm-1*$^{+/+}$*bnaC.fancm-1*$^{+/+}$ or *bnaA.fancm-1*$^{+/+}$*bnaC.fancm-2*$^{+/+}$*)* (green square*)* were obtained for each F1. For some F1 (e1-2v and e1-8v) both F2 (orange line) and allohaploids (purple line) were obtained. In allohaploids, a single pair of homoeologs is represented.

This approach took advantage of the fact that a cross between plants defective for either *BnaA.FANCM* or *BnaC.FANCM* was necessary to produce a loss-of-function *fancm* mutant in *B. napus* (Figure 4). Owing to the high EMS mutation density expected in these plants (see above), we considered it inadvisable to remove the undesirable background mutation load by backcrossing the F1s to WT plants. Instead, we decided to replicate the experiment by using different mutant alleles and, for each replication, different homozygous recessive mutants that we aimed to compare to homozygous wild-type siblings from the same segregating population (Figure 5). We considered that systematic correspondence between homozygous mutants and increased CO frequencies would lend strong support to the hypothesis that FANCM limits CO formation in *B. napus*.

Concretely, we used as primary biological replicates two F1 hybrids combining *bnaA.fancm-1* with either *bnaC.fancm-1* or *bnaC.fancm-2*. We selfed these F1s and identified in the two resulting F2 progenies double homozygous plants for the two mutations as well as wild type siblings (Figure 4). In total, two *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-1*$^{-/-}$ mutants and two *BnaA.FANCM-1*$^{+/+}$*BnaC.FANCM-1*$^{++}$ WTs were identified in the progeny of the first F1 hybrids, while five *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$ mutants and eight *BnaA.FANCM-1*$^{+/+}$*BnaC.FANCM-2*$^{++}$ WTs were identified in the progeny of the second F1 (Figure 5).

Given the mutation load observed within *BnaA.FANCM* and *BnaC.FANCM*, we anticipated that each F1 hybrid contained an extensive set of segregating EMS mutations; as half of these mutations (hereafter referred as to EMS-SNPs) were to remain at the heterozygous stage in the F2s, they provided ways to measure recombination frequencies (Figure 4). Setting out from these premises, we sequenced the two *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-1*$^{-/-}$ double mutants and their two WT siblings in order to identify in one go: 1) EMS-SNPs that we could use as a source of polymorphism for our genetic analysis and 2) pairs of heterozygous intervals shared between mutant and WT F2s that we could use to compare recombination frequencies. The two single mutants *bnaA.fancm-1*$^{-/-}$ and *bnaC.fancm-1*$^{-/-}$ were also sequenced in order to determine the initial phase (coupling or repulsion) of EMS-SNPs; this was essential to compute recombination frequencies correctly by distinguishing WT vs recombinant allelic combinations. The initial plan was also to sequence a second quadruplet (2 mutants and 2 WTs) from the second F1 (*BnaA.FANCM-1*$^{+/-}$*BnaC.FANCM-2*$^{+/-}$) but this idea was later abandoned for lack of time.

Consistent with mutation density within *BnaA.FANCM* and *BnaC.FANCM* (see above), we detected ~65000 segregating EMS-SNPs genome-wide in each F1. ~20 % (14546/65000) of those mutations were found in exons and led to non-synonymous substitutions (including splice variant and non-sense mutations) in a total of 8751 genes (~8% of total gene number). A subset of those targeted genes (912; ~10%) constituted homoeologous pairs (as established in Chalhoub et al., 2014); in most cases (387/456, 85%), the mutations that we found in both copies of a given homoeologous pair were missense muations.

FANCM limits homologous recombination in *Brassica napus*

We converted a subset of EMS-SNPs into Cleaved Amplified Polymorphic Sequences (CAPS) markers that spanned common genetic intervals between pairs of wild type and mutant F2 plants. We then followed the segregation of these markers to assess recombination frequencies in the F3 progenies. Before performing a genome wide assay with KASPR markers, we assessed recombination frequencies in one genetic interval on top of chromosome A01. This interval was shared between two *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-1*$^{-/-}$ mutants and the corresponding WTs as well as between two *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$ and one of the corresponding WTs. Whereas we did not detect any increase in recombination frequencies in the progenies of the *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-1*$^{-/-}$, we observed a marginally significant increase (~32%, Welch's t-test; p-value = 0.011) in the progeny of *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$ (Table 1). This small variation was partly due to the higher-than-expected genetic distance measured in the progeny of WT *BnaA.FANCM-1*$^{++}$ *BnaC.fANCM-2*$^{+/+}$ (22.6 cM instead of 15.7 cM as in the progenies of *BnaA.fancm-1*$^{+/-}$*BnaC.fancm-1*$^{+/-}$). We therefore focused our effort into defining more genetic intervals between wild type and mutant F2 plants for *BnaA.fancm-1 - BnaC.fancm-2.* This procedure was limited by the fact that: 1) we could only use the EMS-SNPs inherited from *bnaA.fancm-1*$^{-/-}$ (thus only half of the mutations present in the F2) and 2) we had to check individually whether these mutations were heterozygous in both wild type and mutant F2 plants i.e., could be used to define common genetic intervals. In the end, three additional intervals were defined on C01, A01 and A05 but we only assessed recombination frequencies on the interval on C01 so far. Consistent with above, we observed a 1.3 fold (36%) increase in recombination frequencies in the progeny of *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$.

However, the smaller size of this interval reduced the statistical power of our test and this difference appeared not significant (p-value = 0.25) (Table 1). However, the fact that a slight increase of CO frequencies was repeatedly across different intervals in the *bnaA.fancm-1*$^{-/-}$ *bnaC.fancm-2*$^{-/-}$ mutant suggests that FANCM has an antiCO effect in *B. napus*.

Table 1: Homologous recombination frequencies in homozygous wild type and mutant for *BnaA.fancm* - *BnaC.fancm*

	Genetic Distance (mut/wt)	
Intervals	***BnaA.fancm-1* - *BnaC.fancm-1***	***BnaA.fancm-1* - *BnaC.fancm-2***
A01_426-A01_913	15.7 (n=96) / 15.7 (n=72)	29.9 (n=137) / 22.6 (n=116)
C01_477-C01_152	NA	7.71 (n=137) / 5.66 (n=116) /

FANCM limits CO formation in *Brassica napus* allohaploids

Unlike other allopolyploid species (like wheat), CO can form between homoeologous chromosomes in *B. napus* allohaploids (AC, n=19; Grandont et al. (2014); this suggests that the recombination intermediates upon which FANCM could potentially act may also exist in these plants. We thus assessed the effect of FANCM on CO formation between homoeologous chromosomes in allohaploid *B. napus*.

We derived allohaploid progenies from the 2 F1 hybrids that we used in our previous analysis as well as from 5 *BnaA.FANCM-1*$^{+/-}$*BnaC.FANCM-1*$^{+/-}$ and 4 *BnaA.FANCM-1*$^{+/-}$ *BnaC.FANCM-2*$^{+/-}$ additional F1 hybrids (Figure 5). In each of these progenies, multiple pairs of homozygous *bnaA.fancm*$^{-}$ *bnaC.fancm*$^{-}$ mutant and *BnaA.FANCM*$^{+}$ *BnaC.FANCM*$^{+}$ wild type were recovered and used to compare homoeologous recombination frequencies using cytological approaches. This assay therefore encompassed two layers of replications: 1) the F1 hybrids that we used to derive allohaploids and 2) the different *bnaA.fancm*$^{-}$ *bnaC.fancm*$^{-}$ mutant and *BnaA.FANCM*$^{+}$ *BnaC.FANCM*$^{+}$ wild type that were derived from a given hybrid. We considered that systematic correspondence between *bnaA.fancm*$^{-}$ *bnaC.fancm*$^{-}$ mutants and increased CO frequencies across all F1s and all allohaploids was necessary to support the hypothesis that FANCM limits CO formation between homoeologous chromosomes in *B. napus*.

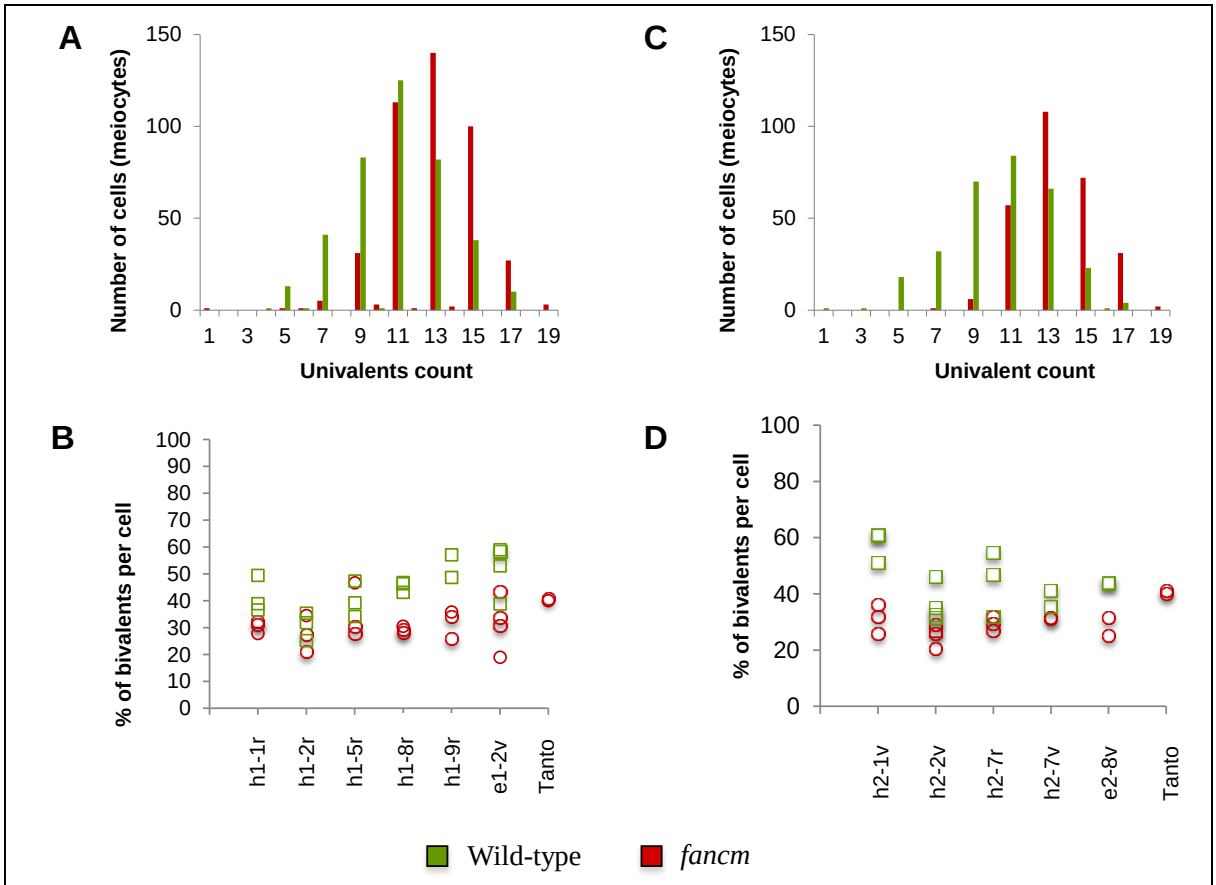


Figure 6: Homoeologous recombination in allohaploids wild-type and *fancm* plants

Homoeologous recombination is assessed in *bnaA.fancm-1*$^{-}$ in combination with either *bnaC.fancm-1*$^{-}$ (A, B) or *bnaC.fancm-1*$^{-}$ (C, D). At least 3 allohaploids mutants (*bnaA.fancm-1*$^{-}$*bnaC.fancm-1*$^{-}$ or *bnaA.fancm-1*$^{-}$ *bnaC.fancm-2*$^{-}$*)* and 3 wild types *(BnaA.FANCM-1*$^{+}$*BnaC. FANCM -1*$^{+}$ or *BnaA. FANCM -1*$^{+}$*BnaC. FANCM -2*$^{+}$*)* were derived from 6 (A, B) and 5 (C, D) heterozygous F1 plants for both BnaA.*FANCM* and BnaC.*FANCM*. The number of univalent between mutants and wild types allohaploids has been obtained when pooling all biological replicates together (A, C). The number of bivalent per cell between mutants and wild types is reported according to the heterozygous plants used to produce the allohaploids (B, D). Meiotic recombination has been observed in 20 meiocytes per plant.

When comparing homozygous mutant and wild type plants for *BnaA.fancm-1* - *BnaC.fancm-1*, we observed a marginally significant increase in recombination frequencies, i.e., decrease in the number of univalents, in *fancm* mutants (mean number of univalents = 12,8 in wild type compared to 10,9 in *fancm* mutant, Wilcoxon signed rank test, p-value = 0.02582). However, this increase in recombination frequencies was not consistent across the 6 F1 hybrids (Figure 6A and 6B). For three of them, the number of univalent was essentially the same between WT and *fancm* allohaploids. One of these F1 was e1-2v, for which we did not observe any increase in homologous recombination frequencies in the progenies of F2 *fancm* mutants (Figure 5, Table 1).

In contrast, we observed a significant and consistent increase in recombination frequencies across all 5 hybrids when comparing homozygous mutant and wild type plants for *BnaA.fancm-1* - *BnaC.fancm-2* (mean number of univalent = 13,5 in wild type compared to 10,5 in *fancm*, Wilcoxon signed rank test, p-value = 0.001662). This trend was observed for all allohaploids and all F1 hybrids, with some variation in the magnitude but no variation in the direction of the change (Figure 6C and 6D). In the case of e1-8v, which was used for assessing homologous recombination rate (Figure 5, Table 1), the increase in recombination frequencies between WT and *fancm* allohaploids (mean number of univalent in *fancm* = 10.6 compared to 13.6 in wild type) exactly matched the mean difference measured across all samples.

***BnaC. fancm -2 is* most likely non-null**

Given the small but significant increase of CO frequencies repeatedly observed in *bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$, we assessed the extent to which the substitution identified in *bnaC.fancm-2* is detrimental for FANCM anti-CO activity. In order to do so, we transformed an *A. thaliana msh5 fancm* double-mutant with a modified copy of *At.FANCM* mimicking *BnaC.fancm-2*. The *A. thaliana msh5 fancm* double-mutant is fertile, because FANCM deficiency restores bivalent formation in the *msh5* CO-defective mutant (Crismani et al., 2012); we reasoned that the transformant should remain fertile if *BnaC.fancm-2* leads to a completely non-functional protein, while partial or complete sterility should be restored if BnaC.FANCM-2 is functional. We observed that the triple mutants were partially sterile, suggesting that BnaC.FANCM-2 is still functional. Cytological analysis (bivalent - univalent counts) are ongoing in the triple mutants i) to check that the partial sterility that we observed is due a defect in meiosis, ii) to assess the extent to which bivalent formation is restored in the triple mutant compared to a single *msh5* mutant.

Discussion

FANCM limits recombination frequencies within the *Brassica*

Altogether our results indicate that the anti-CO activity of FANCM is conserved in two important *Brassica* crops, thus possibly across the entire *Brassicaceae* family.

This point is more strikingly illustrated in *B. rapa* where we observed a 2.5-fold increase of bivalents in the *fancm/msh4* double mutant compared with the single *msh4*. This change is consistent with the 3-fold increase of COs reported in *A. thaliana* (Crismani et al., 2012); alike Arabidopsis, the extra COs were sufficient to restore bivalent formation to a wild-type level in *B. rapa*. A lesser pronounced increase of CO frequencies (~1.3 fold) was observed in *B. napus*. Although marginally significant from a statistical point of view, the same increase was repeatedly observed across two independent genetic intervals in euploids (Table 1) and across all biological replicates in allohaploids (Figure 6D) produced from *BnaA.FANCM-1*$^{+/-}$ *BnaC.FANCM-2*$^{+/-}$. This is unlikely to be a mere coincidence, especially in view of the results obtained in the euploid and allohaploid progenies of *BnaA.FANCM-1*$^{+/-}$*BnaC.FANCM-1*$^{+/-}$. Instead, these results lend support to the hypothesis that FANCM limits CO formation in *Brassica* (crop) species.

This interpretation is consistent with the high level of identity shared between *Brassica* and *A. thaliana* FANCM proteins (Figure S1), in particular in the bipartite helicase domain where all mutations were identified (in *A. thaliana, B. rapa* and *B. napus*). The even higher protein sequence identity shared between Brassica FANCM homologs does not suggest that FANCM activity might vary extensively between these species.

The most straightforward hypothesis to explain the discrepancies observed in the magnitude of CO increase between *B. rapa* and *B. napus* (see above) is that *BnaC.fancm-2* is not a complete loss of function mutation. Residual anti-CO activity of *BnaC.fancm-2* is consistent with the partial sterility observed in the triple *BnaC.fancm-2-msh5-fancm A. thaliana* mutant (although it remains to be established that this is due to a shortage of bivalent formation). It is not known however if *BnaC.fancm-2* is over-expressed in this transformant (as compared to *bnaA.fancm-1$^{-/-}$bnaC.fancm-2$^{-/-}$*), which could limit the extent to which strong conclusion can be drawn from this experiment. In addition, integrating the small increase of COs observed in *B. napus fancm* mutants with a possible residual anti-CO activity of *BnaC.fancm-2* leads to question the extent to which FANCM anti-CO activity is dosage dependent; i.e., whether a "leaky" *FANCM* allele could result in intermediate CO increase.

In the absence of a confirmed loss of function mutant (induced stop) in BnaC.FANCM, this remains an open question.

Alternatively, part of the discrepancies observed in the magnitude of CO increase between *B. rapa* and *B. napus* could result from a positional effect of the tested genetic intervals and/or the inhibitory effect of EMS-SNP heterozygosity on extra CO formation in the *B. napus* mutants. In *A. thaliana*, the impact of FANCM on CO recombination is not homogenous across the genome and can vary extensively between close intervals within the same region (from ~2 fold to ~3.5 fold in adjacent interval on chromosome 3) (Crismani et al., 2012). In this study, I selected intervals on top of chromosome A01 and C01, where recombination is expected to be the highest (See paragraph 4.2, p.61); these intervals were thus the best place to detect CO variation. The anti-CO activity of FANCM was shown to be negligible when assessed in *A.thaliana* hybrid progenies compared to pure lines (Girard et al., 2015); likewise, in recombinant plants, juxtaposition of heterozygous and homozygous regions was shown to drive CO inhibition and CO promotion, respectively (Ziolkowski et al., 2015). It is thus reasonable to think that the level of sequence divergence between progenitors determines the extent to which FANCM deficiency can cause CO increase. In our assay, we used highly inbred plants, where the only polymorphisms were introduced by the EMS treatment. However, the observed EMS-SNP density is quite low (~65000 EMS-SNPs per plant; ~ 1 SNP every 18Kb) compared to the SNP density in *A.thaliana* (1 SNP every ~200pb, Crismani et al., (2012) and *B. napus* (~ 1 SNP every 1,5kb in cultivated varieties; Trick et al. (2009) hybrids. We cannot exclude however that this intermediate level of heterozygoty would have somehow reduced, but not completely abolished the anti-CO activity of FANCM in our mutant plant.

Does FANCM limit homoeologous recombination frequencies in *Brassica napus?*

The preceding discussion on the consequences of heterozygosity on extra CO formation in *fancm* hybrids casts doubts upon the effect of FANCM on homoeologous recombination in *B. napus* allohaploids. The polymorphism rate between the A and the C sub-genomes of *B. napus* is much (~30-fold) higher than the SNP density measured between different varieties (i.e. between homologous regions). For example, focusing on gene models, we estimated that A and C homoeologous copies showed an average of ~2 % divergence, which is consistent with the ~ 3,5% divergence measured within transcripts by (Cheung et al., 2009). If the SNP density observed in Arabidopsis (0.5%) is sufficient to inhibit CO formation in *fancm* mutants, it is surprising that a higher amount of variation would allow some extra COs to mature in *B. napus* allohaploids. Yet we observed a 1.3 to 1.8-fold increase in bivalents formation in homozygous mutant for *BnaA.fancm-1* - *BnaC.fancm-2*, which is consistent with the increase in CO frequencies measured between homologous chromosomes in euploids (AACC) *fancm* mutants. As described in paragraph 4.2 (p.61), some homologous regions are shared between homoeologous chromosomes as a consequence of homoeologous exchanges (HE). Although the HE landscape has not been characterized yet in cv. *Tanto* (Chalhoub et al., 2014), there is no doubt that this variety contains at least some HEs. It is therefore tempting to hypothesize that the increase in bivalent formation observed in the allohaploids AC *fancm* mutants results from an increase of CO formation within the homologous regions that were duplicated on homoeologous chromosomes by HE fixation. On that assumption, the difference observed between *fancm* mutants and WT allohaploids would reflect a difference of homologous rather than homoeologous recombination. Testing this hypothesis is however not straightforward, as it would require assessing (i) whether the increase in CO frequencies occurs in specific chromosomal regions and (ii) whether these regions co-localize with an existing HE.

All of this aside, our results on allohaploids support the interpretation that FANCM limits CO frequencies in *B. napus*, even if the extent to which CO rate can be increased still remains to be established.

The hurdles of translational biology

Without limiting the foregoing, the scope of the genetic assay has been somehow reduced as we were not able to repeat the analysis by using different mutant alleles and because we assessed CO frequencies on only two genetic intervals. This originates from the fact that we did not find a non-sense mutation for *BnaC.FANCM* and because sequence identity was of little use to predict amino acid substitutions deleterious for FANCM anti-CO activity.

Retrospectively, it is possible to estimate that the odds not to find any "STOP mutation" in ~500 M2 plants of the mutagenized population that we used were not negligible (~25%). In the absence of a "STOP mutation", we chose nonsense mutations targeting highly conserved amino acid in the helicase domain of FANCM. However, amino acid conservation in this case was of limited help to assess the effect of the substitution. The case of *BnaC.fancm-2* in that respect is quite illustrative; although it targeted a highly conserved amino acid in the helicase domain of FANCM, whose substitution is causal for a loss of function FANCM in *A. thaliana*, we highly suspect that the resulting protein is still active in *B. napus*. As our experimental design calls for replicate, it might be worth adjusting the TILLING strategy accordingly to increase the chance to recover more than one loss of function mutant. A higher throughput approach, like TILLING by sequencing (Tsai et al., 2011) would be a strategy worth exploring.

However, it remains true that TILLING always relies on EMS mutations with many off-targets. In our assays, the mutation density in the allotetraploid *B. napus (*1 EMS-SNP every 12-14Kb) was twice the mutation density found in the diploid *B. rapa (*1 EMS-SNP every 31Kb). High mutations density have been repeatedly observed in EMS mutagenized population in polyploids where the presence of multiple copies for a gene is thought to buffer the effect of deleterious mutations (Slade et al., 2005; Wang et al., 2008). This entails a risk of background EMS mutations being mistaken for mutations in target genes. Classically, backcrossing to WT is recommended to purge the background mutations off the targeted mutant. This is not possible in practice in *B. napus* where after 20 backcrosses we would still have to count with ~1000 EMS-SNPs segregating in the genome. The risk of confusion between background mutations and mutant allele is however mitigated as we looked at a very specific phenotype (reduction of CO frequencies) for which only a few genes are known to contribute.

Furthermore, most of these genes are present in multiple copies in *B. napus* (Blary, Lloyd et al., in prep) that could provide a buffering effect against the impact of non-target EMS mutations. It is possible to estimate that the odds to find putative defective off target mutations in all copies of a gene encoding an anti-CO protein other than FANCM is one every ~19000 lines. Thus, although we cannot completely discard the risk of confounding the effect of an off-target mutation with FANCM, this remains very unlikely.

Conclusion

Gaining control over CO patterning in crops could be instrumental for plant breeding. In this study, we showed that it is possible to increase CO frequencies in *B. napus* and *B. rapa*, two representative crops within the *Brassica* genera by mutating FANCM, an anti-CO protein. Although further studies are needed to evaluate precisely the extent to which CO frequencies can be increased in *B. napus*, this work illustrates how translational biology can open the way to novel applied possibility.

The obtention of FANCM mutants in *B. napus* and *B. rapa* was instrumental to assess whether it could be possible to further increase CO frequencies in *Brassica* triploids hybrids (see paragraph 2.4, p.56). Although it was one of the objectives of my PhD (as explained in Chapter 3, p.57), I was not able to produce the right triploid *Brassica* mutant for FANCM. Because the TILIING experiment has been performed in incompatible *B. rapa* and *B. napus* genotype, I only managed to produce some triple mutants *bnaA.fancm-1*$^{-/-}$*_braA.fancm-1*$^{-/-}$*_bnaC.fancm-1*$^{-/-}$ but unfortunaltely not *bnaA.fancm-1*$^{-/-}$*_braA.fancm-1*$^{-/-}$*_bnaC.fancm-2*$^{-/-}$. As *bnaC.fancm-1* is not a loss of function mutation, we did not expect to observe any difference in CO frequencies between wild-type and *bnaA.fancm-1*$^{-/-}$*_braA.fancm-1*$^{-/-}$*_bnaC.fancm-1*$^{-/-}$. Therefore, this part of the project was discontinued.

Material and methods

FANCM protein identification in *Brassica* – screening of the BAC libraries

Homologues and putative homologues of FANCM were identified using literature searches and reciprocal BLASTp and PSI-BLAST (http: //www.ncbi.nlm.nih.gov/. The Screening of the *Brassica napus* BAC "DarmorBZH" library was performed by the CNRGV (INRA Toulouse).

Plant material

The genealogy of the plant material used in this study is detailed in Figure 5. The plants were grown in standard long day greenhouse conditions. The allohaploids plants were obtained following the protocol described in (Jenczewski et al., 2003).

TILLING experiment

We looked for mutation in *BraA.FANCM* in the EMS mutant population of *B. rapa subsp. trilocularis* (Yellow Sarson) developed by the John Innes Centre, RevGenUk (Stephenson et al., 2010). The following primers were used to amplify a region of 1187bp in *BraA.FANCM:* Bra034416_F1 3'-TGGCAAGGGATAAGTTTCGTGAAGCAC-5' and Bra034416_R2 5'-GGCATAATCCGATAAAAGTGGCACTGG-3'.

We looked for mutation in *BnaA.FANCM* and for *BnaC.FANCM* in the population of *B. napus*, genotype *Tanto* developed by Nathalie Nesi at INRA Rennes. Primers were designed to amplify a single locus and tested to ensure that only one of the homoeologous loci was amplified. The following primers were used to amplify a region of 1131bp in *BnaA.FANCM:* T_FANCMAF1 5'-CCAAAATGTGTTCCAAATTCATC-3' and T_FANCMAR2 5'-GGGATGGTTTAAGAACAAATCATA-3'. The following primers were used to amplify a region of 1129bp in *BnaC.FANCM:* T_FANCMCF1 5'-CCAAAATGTGTTCCAAATTCATT-3' and T_FANCMCR2 5'-GGGATGGTTTAAAAACAAATCAAG-3'.

Genotyping of *BnaA.fancm -1, BnaC.fancm -1, BnaC.fancm -2*

In *Brassica napus,* the genotyping of *bnaA.fancm -1* was performed using T_FANCMAF1-T_FANCMAR2 followed by a digestion with XmnI/PdmI (698+433 for the wild type amplicon and 1129bp for the mutant). The genotyping of *bnaC.fancm -1* was performed using T_FANCMCF1- T_FANCMCR2 followed by a digestion with DraI (1129bp for the wild type amplicon and 972+157 for the mutant). The genotyping of *bnaC.fancm -2* was performed using T_FANCMCF1- T_FANCMCR2 followed by a nested amplification using the following

primers pairs: dcapsFANCMC1F 5'-CATTCGCAAGCTTCTTCCTAGTCAT-3' and dcapsFANCMC1R 5'- TTGGACAATTTCGGGCTTGG-3' and a digestion with BsiYI (215+23 for the wild type amplicon and 238 for the mutant).

DNA extraction – Sequencing- EMS-SNP detection

Total DNA were extracted using the NucleoSpin® Plant II Midi / Maxi (Macherey-Nagel) extraction kit. The DNA sequencing was carried out at the Institute of Plant Sciences Paris-Saclay (IPS2, Saclay, France). Both the single homozygous mutants for *bnaA.fancm-1* and *bnaC.fancm-1* were sequenced on the same single line of an Illumina HiSeq sequencing system. The corresponding double homozygous and wild type were sequenced on a line each. Mutations were identified thought MutDetect pipeline developed by Bioinformatic and Informatics IJPB team (Girard et al., 2014).

Genetic assay to measure CO frequencies between homologous chromosomes

Heterozygous EMS mutations were converted in copy specific CAPS or dCAPS primers when shared between at least one pair of homozygous wild type and mutant for *FANCM*. The list of the primers used in this study is given in Table S1. Recombination frequencies were estimated using MapDisto (Lorieux, 2012).

Cytology techniques

Briefly, the meiotic behaviour was observed on pollen mother cells at metaphase I. To assess CO frequencies between homologous chromosomes, male meiotic spreads for DAPI staining were prepared as described by Chelysheva et al. (2013) from buds fixed in Carnoy's fixative (absolute ethanol:acetic acid, 3:1, v/v).

To assess CO frequencies between homoeologous chromosomes, 20 pollen mother cells were examined in each allohaploid to obtain the mean number of univalent, bivalents or multivalent per cell.

Pyrosequencing

Pyrosequencing was performed on meiotic cDNA and on gDNA to check for amplification bias. The following primers were used for amplification and sequencing:
pFANCMR:TTTCGTTGGCTAAATCTTCTTCCT,
pFANCMF:ACGAAGCAAACAGAGAAGAAGACC,
pFANCMS:TCTTCTGCCAATTCATTA

Primer pairs have been designed with Pyromark Assay Design v2.0.1.15 and the

pyrosequencing reaction has been performed with PyroMark Q24 v2.0.6 of QIAGEN®.

Directed Mutagenesis Constructs, Plant Transformation, and Plasmid Constructs

Amplification of *FANCM* genomic fragment covered 618 nucleotides before the ATG and 1029 after the stop codon. The PCR product was cloned, by Gateway (Invitrogen) into the pDONR207 (Invitrogen) to create pENTR-FANCM, on which directed mutagenesis was performed using the Stratagene Quick-change Site-Directed Mutagenesis Kit. For plant transformation, LR reaction was per-formed with the binary vector pGWB1 (Nakagawa et al., 2007). The resulting binary vectors were trans-formed using theAgrobacterium-mediated floral dip method (Clough and Bent, 1998) on double homozygous mutant plant (*fancm*$^{-/-}$/*msh5*$^{-/-}$).

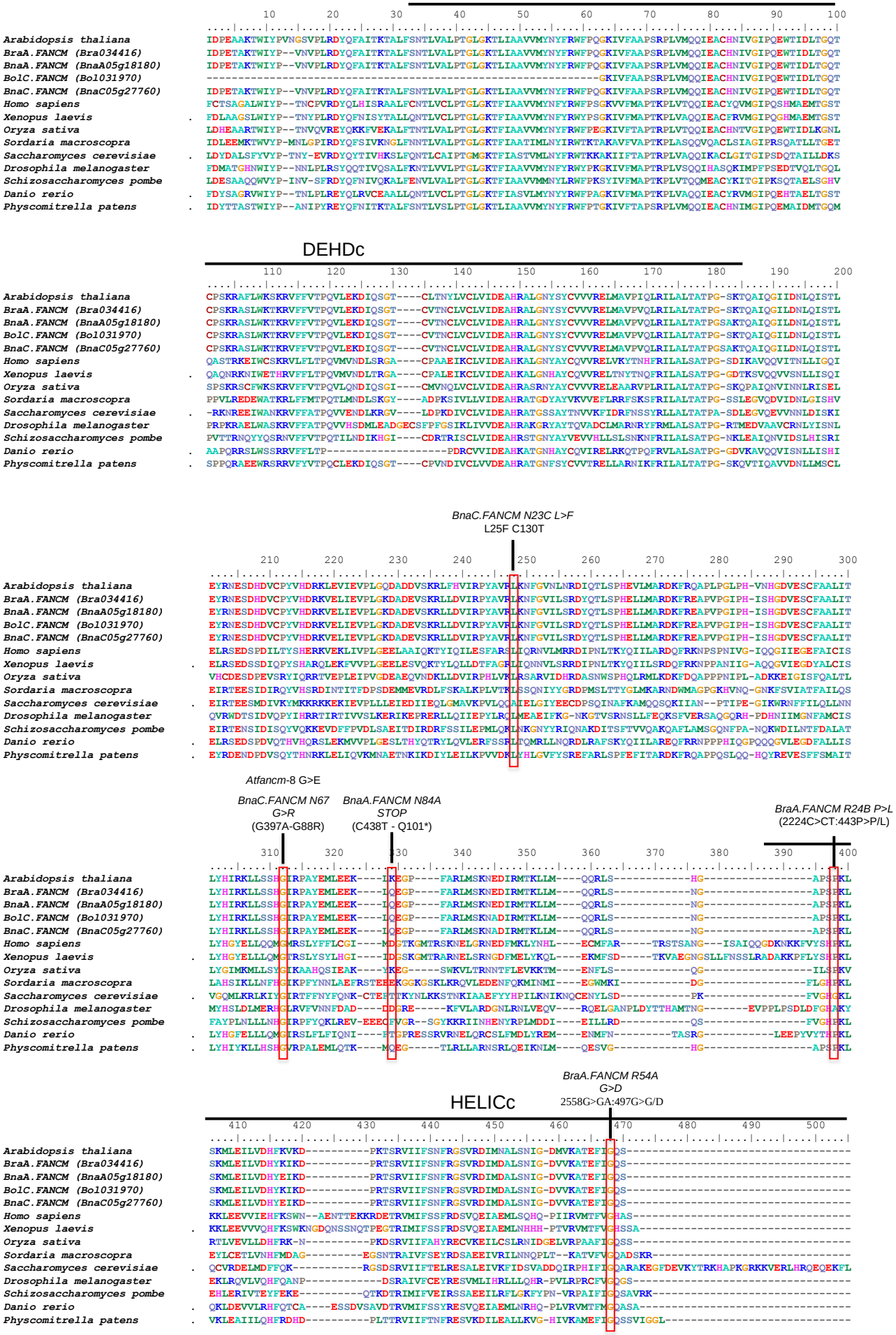
Annexes

Supplementary Table

Table S1: List of primer pairs used in the genetic assay to measure CO frequencies between homologous chromosomes

Chr. No.	Coordinates	Enzyme	Forward primer sequence	Reverse primer sequence
A01	77426	FokI (85/50/35)	TTGCATTGGTCCACACCCCAAGGAT	CAGGTTCAGCTTGCTCACAGGTGG
A01	1688913	BseNI (514/371/143)	ATGGGAACGACAGGACTGAG	TGTCCTCACCATCGGCTAAA
A02	23312988	HaeIII (396/372/24)	TTCTATTATTCAAATTAGAGATGGC	AGTAACATTGTGTGAGATTTGTCT
A02	24768015	AciI (70/47/23)	AAGAGAATGTAGAAAGTGTTGG**C**CG	TACTTTCTCACTTCCACCCACCACA
A05	1012215	BsiYI (66/44/22)	ACGACATCCAGTTTGCAGAT	GTGAAGTGGAAAATTTTAAGCATGA
A05	2147252	MjaIV (364/300/188/92/64)	TTTTCTTACATGATCCTCCAGAAG**G**	CTCGTTTGTGTCTAGAAGCTTTTC
C01	81477	AciI (74/51/23)	AAACACGAAAATTTAGAGAAC**C**G	GTTTCAACGCTTCCCAATGC
C01	657152	SfaNI (198/173/134/64)	CCAATGGGGTTTAATGGGCTC	AGACTCGAAAGGTTCCAGCA

Supplementary Figures


DEHDc
BnaC.FANCM N23C L>F
L25F C130T
Atfancm-8 G>E
BnaC.FANCM N67 G>R
(G397A-G88R)
BnaA.FANCM N84A STOP
(C438T - Q101*)
BraA.FANCM R24B P>L
(2224C>CT:443P>P/L)
HELICc
BraA.FANCM R54A G>D
2558G>GA:497G>G/D
Arabidopsis thaliana
BraA.FANCM (Bra034416)
BnaA.FANCM (BnaA05g18180)
BolC.FANCM (Bol031970)
BnaC.FANCM (BnaC05g27760)
Homo sapiens
Xenopus laevis
Oryza sativa
Sordaria macroscopra
Saccharomyces cerevisiae
Drosophila melanogaster
Schizosaccharomyces pombe
Danio rerio
Physcomitrella patens

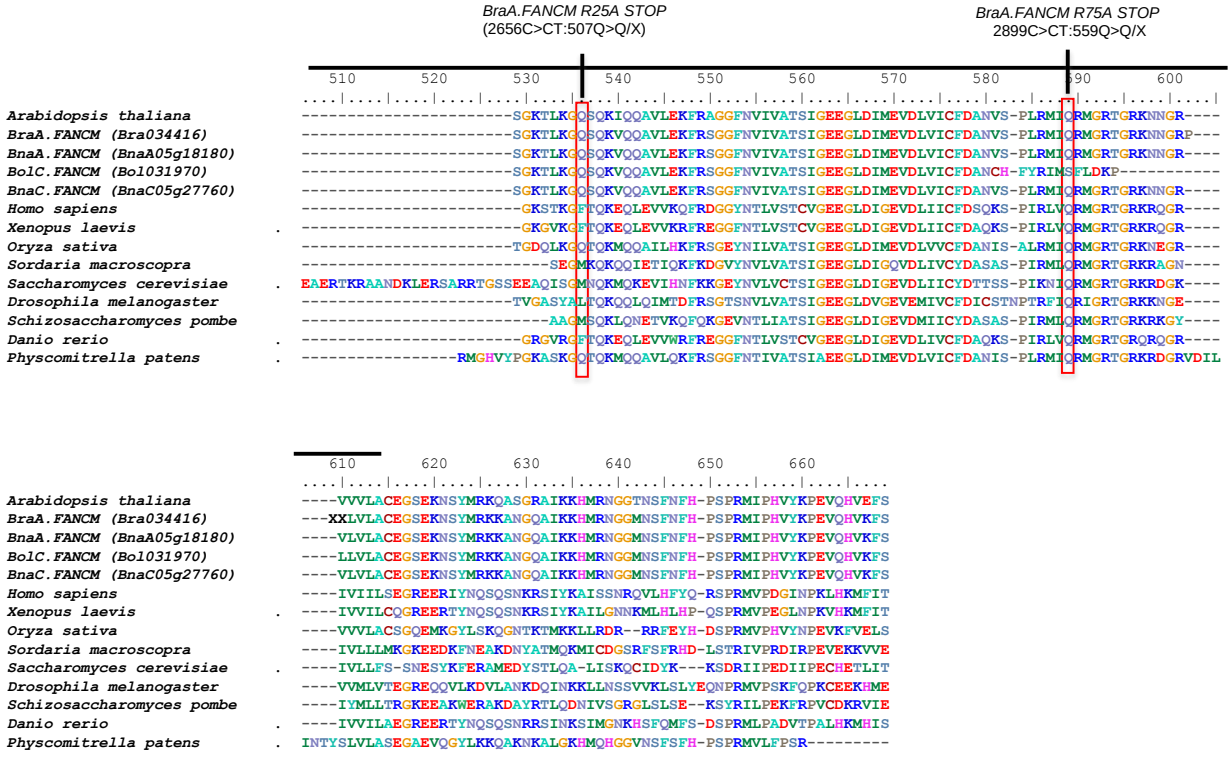


Figure S1: ClustalW multiple alignment of the helicase region of FANCM in *Brassica* and other species representative for a family

The positions of the mutations identified in this study are shown along the N-terminal bipartite SF2 helicase domain of FANCM. Arabidopsis thaliana NM_001198212;; Hs FANCM NP_065988.1; Xl NP_001171151.1; Os AAX96303.1; Sordaria macrospora XP_003348274.1 ; Sc Mph1 NP_012267.1, Dm NP_650971.2; Sp Fml1 Q9UT23.2 ; Dr NP_001107132.1; Pp XP_001753469.1;

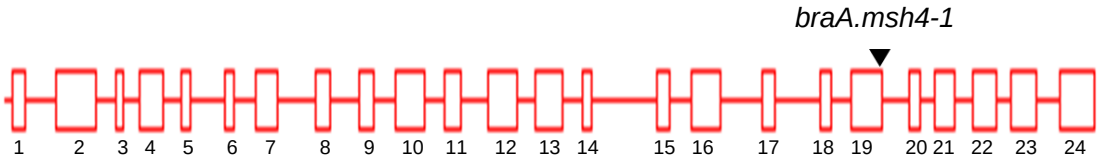


Figure S2: Position of *braA.msh4-1*

A black arrow points to the position of the mutation *braA.msh4-1* on *BraA.MSH4*. *braA.msh4-1* induces a substitution in the acceptor site of the 19th exon; the composition intron (line) exon (box) of *BraA.MSH4* is given. *BrA.MSH4* coding sequence contains 794 aminoacids.

Chapter 6 General discussion

6.1 The meiotic transcriptome is highly variable within *Brassica napus*

Our RNA-Seq dataset indicates that more than 45% of the *B. napus* gene set is expressed in meiocytes. This is likely an under-estimation as many un-annotated features were also found to be transcribed during meiosis in *B. napus*, as in other species (Dukowic-Schulze et al., 2014a; Flórez-Zapata et al., 2014). These observations raised the question of the regulation of the meiotic transcriptome.

To address this issue, a comparison was undertaken between the meiotic transcriptome of two representative genotypes of *B. napus*. Altogether our results show that these meiotic transcriptomes vary extensively, if one uses the number of genes differentially expressed as an indicator. Among the three factors (genome, genotype and ploidy) that we controlled during this analysis, the difference in expression between homoeologous gene pairs (genome effect) contributed the most to the variations observed, followed in order of importance by the genotype (*Darmor-bzh* and *Yudal)* and the ploidy effect (allohaploid AC and euploid AACC). However, although these differences were consistent across all replicates (hence their high level of statistical significance), most of them were of limited amplitude. These very limited fold-change prompt questions as to the biological relevance of these changes. Although small difference in expression can cause severe phenotype (Ruzycki et al., 2015), transcript abundance is usually only a poor predictor for the final quantity the corresponding protein (because of post transcriptional and post traduction regulation) [reviewed in (Rose et al., 2004; Alós et al., 2008)]. It is thus likely that most of these variations did not have any impact on the meiotic phenotype. This does not mean however that some of them could not drive change in the meiotic behavior.

Our analysis indicated that differential expression between A or the C copies within homoeologous gene pairs accounted for the highest number of genes showing transcriptional variation in our dataset. There was however no evidence that one sub-genome contributed more than the other to the total transcriptome.

Large difference in expression within *B. napus* homoeologous gene pairs is not specific to the meiotic transcriptome; only ~58% of homoeologs were found to contribute similarly to gene expression in leaves and roots (Chalhoub et al., 2014). It is noteworthy that for ~4% of homoeologous gene pairs, an higher expression of the A compared to the C copy was found in leaves, while the reverse was true in roots (Chalhoub et al., 2014). This is an interesting information in the context of HE. For one of these genes, an event ($A_n^+ \rightarrow C_n^-$) will not have the same consequences on the total (A+C) expression in leaves and in root (total A+C expression will be higher in leaves).

The second factor that contributed to most of the total genes showing transcriptional variation was the genotype effect (*Darmor-bzh* vs *Yudal*). We indeed observed that more than 60% of the genes that were expressed in meiocytes were differentially expressed between *Darmor-bzh* and *Yudal*. This trend holds true for meiotic genes and the transcription factors that were consistently overexpressed in *A. thaliana* and maize meiocytes compared to somatic tissues (see below).

The extent to which the meiotic transcriptome varies between *Darmor-bzh* and *Yudal* came as a surprise to us, in part because only a few studies have compared meiotic transcriptome across different genotypes. Basically, I am aware of only two studies addressing this issue. When comparing different yeast strains that differ in sporulation properties, Primig et al., (2000) found a subset of approximately 900 core genes over 1600 meiotically regulated genes that displayed a strain-independent pattern of meiotic transcriptional regulation. Recently, Flórez-Zapata et al., (2016) compared the meiotic transcriptome of prophase I meiocytes extracted from different sunflower genotypes that differ in CO frequencies and observed ~50% of differentially expressed genes. Our observation is roughly consistent with these estimates.

We can then ask ourselves about the causes of such transcriptional variation between genotypes. Two main hypotheses can be envisaged.

First, although we took every precaution to control that RNA extractions were performed on meiocytes sampled at the same meiotic stage in *Darmor-bzh* and *Yudal*, we cannot completely rule out the hypothesis that this was not exactly the case. *Darmor-bzh* and *Yudal* euploids do indeed differ in the progression of meiotic recombination (Grandont et al., 2014). Under the hypothesis that meiotic gene expression occurred in successive waves that are each associated with specific stages and/or events of meiosis (Crismani et al., 2006), the differences that we observed could only reflect the fact that we are comparing snapshot of meiotic transcriptomes that are slightly shifted in time. When looking specifically at meiotic genes, we found no evidence for preferential expression of genes that are thought to act early or late during meiosis between *Darmor-bzh* and *Yudal*. In *A.thaliana*, DUET which is likely to function as a positive regulator of gene expression, is only expressed at the diplotene stage (Andreuzza et al., 2015). In *B.napus*, we found expression for DUET in both *Darmor-bzh* and *Yudal*, although slightly overexpressed in *Yudal*. Although we did not found evidence pointing towards a confounding effect of the timing of meiosis, this remains an open question as there is no way to test whether prophase substages have the same duration in *Darmor-bzh* compared to *Yudal*.

Second, the differences in expression of a large number of genes could result from a domino effect due to the differential expression of transcription factors in *Darmor-bzh* and *Yudal*. In Dukowic-Schulze et al., (2014), the authors found a common subset of transcription factors overexpressed in meiotic compared to mitotic tissue in both *A. thaliana* and *maize*. Interestingly, we found that these transcription factors, as well as numerous meiotic genes, were clearly overexpressed in *Yudal*. It is tempting to imagine that the differential expression of key transcription factors could ultimately result in massive differential expression for a large number of genes. As the regulatory pathway for those meiosis specific transcriptions factors have not been characterized yet, this hypothesis remains tentative.

Finally, it must be acknowledged that these two hypothesis are far from excluding each other. On the contrary, one may envisage that a differential timing in the progression of meiosis could lead to activation of different transcription factors and thus result in massive number of slightly differentially expressed genes.

In contrast to the unexpected large difference in gene expression observed between *Darmor-bzh* and *Yudal*, we found little variation between allohaploids (AC) and euploids (AACC) for a given genotype. Although a fair number of genes were differentially expressed between plants with different ploidy, less than 1% of them showed a log2 (fold change) >1.
In view of the variation discussed above, and given what I presented in paragraph 1.3, p17, this result appears counter intuitive. Despite the tight integration between meiotic recombination and cell cycle progression, it is surprising that little change in gene expression was found in allohaploids where meiosis is obviously disturbed. However, the first apparent meiotic defect in allohaploids are observed at pachytene (uncomplete synapsis; Grandont et al., 2014), which take place >16 hours after the S phase in Arabidopsis (Armstrong et al., 2003; Stronghill et al., 2014). It is thus possible that the negative feedback loops that tie DSB formation to SC formation has not been triggered in the meiocytes that we used to extract RNA, simply because pachytene cells may be under-represented in this sample.

6.2 Phenotypic consequences of HEs

The analysis of the meiotic transcriptome of *B. napus* revealed an unexpected source of variation both between *Darmor-bzh* and *Yudal* but also between *Darmor-bzh* biological replicates: i.e. the presence of differentially fixed or still segregating homoeologous exchanges (HEs). I showed in paragraph 4.2 (p.61) that these HEs had a significant impact on both gene content and gene expression. I have thus treated this source of variation separately to avoid confounding effect.

We first provided a solid ground for our transcriptome analysis by confirming that the HE content was identical between our genotypes (*Darmor-bzh* and *Yudal*) and the lines analyzed by Chalhoub et al. (2014). This revealed to be a necessary step as we did not confirm all the HEs described in Chalhoub et al. (2014): only 15/17 and 12/13 HEs were validated in *Darmor-bzh* and *Yudal*, respectively. We observed that HEs were preferentially located in the most distal third of chromosome arms where CO frequencies is high. HEs are also located in regions of high gene density; we estimated that the HEs fixed either in *Darmor-bzh* or in *Yudal* together encompass a few thousand (> 3,500) gene models. Because of the resulting difference in gene content between these two genotypes, all HEs generated divergent gene expression profiles.

Both the duplication and the concurrent loss of a gene as a result of an HE resulted in clusters of highly differentially expressed genes that could be used to detect HEs without any prior indication of their position.

We then characterized a newly-formed HE that we found segregating among three *Darmor-bzh* biological replicates. This event led to the loss of one (ACCC) or two (CCCC) copies of the A sub-genome in a single chromosomal region at the top of A_n1-C_n1. In contrast to the fixed event that are fixed in *Darmor-bzh* or *Yudal,* this HE encompassed a very large region (4.4 Mb or 1470 genes, some of which are specific to the A region). This segregating event offered a unique opportunity to evaluate the extent to which variation in gene copy number correlates with gene expression change. Overall we observed that the level of expression of a gene in the newly formed HE was directly proportional to the number of copies of that gene. Thus duplication of a dominantly expressed homoeologue prior to HE leads to increased $Total_{(A+C)}$ expression while duplication of a lesser expressed homoeologue results in reduced $Total_{(A+C)}$ expression.

In older events, i.e., fixed HEs, absolute dose difference did not appear to be the only determinant of gene expression, transcriptional compensation and/or selection against HEs might have occurred to re-establish expression levels prior to HE.

The phenotypic consequences of structural variation (SV) is considerable in human populations where large scale genomic alterations have been associated with common and rare human disease (Weischenfeldt et al., 2013a). Although SV has been hypothesized to be a driving force behind phenotypic variation in plant (Chia et al., 2012), their study has been more limited. Substantial progress has however been made over the last few years as genome wide detection of SV has been made possible through recent technological advances in genome sequencing. Extensive resequencing of genotypes within a species has allowed the capture of the substantial amount of variation that lie outside the single reference genome (pangenome), not only in the form of SNPs and small indels but also in the form of large scale genomic alterations that can result in intraspecific differential gene content (Wendel et al., 2016). In maize for example, dispensable genes i.e., the acknowledged set of genes that is only present in some but not all individuals of a species (Albalat and Cañestro, 2016) represent at least 50% of the genome.

Structural variation accounts for a part of this intraspecific variability in gene content. This source of diversity can be of interest to breeders. Genome Wide Association Studies (GWAS) in maize have found that SNPs within PAV, i.e., sequences that are present in one individual but absent in another (Springer et al., 2009), are enriched for significant GWAS hits for agronomic traits compared to the fraction of SNPs outside PAV (Lu et al., 2015). In other crop species, structural variation has been associated with diverse traits like biotic and abiotic resistances (nematode cyst in soybean, and aluminium and boron tolerance in barley) or reproductive morphology in cucumber (Zhang et al., 2015) [see Saxena et al. (2014) for review].

In this study, we have shown that segregation of large homoeologous exchanges (HEs) contributes to differential gene content between allopolyploid genotypes. HE formation is driven by crossover formation between homoeologous chromosomes and results in the replacement of one chromosomal region (which is lost) with a duplicate of the corresponding homoeologous region. As such, HEs are distinct from mere gene copy number variation. The biological consequences of such events comes down to asking whether homoeologues are able to compensate for one another. We have already seen in paragraph 4.2 (p.61) that this is not necessarily the case. Briefly we obtained evidence that support the belief that HEs can be selected against when duplication of lowly expressed genes results in detrimental expression level. Thus expression divergence between homoeologs prior to HEs as well as differences in gene content are likely to result in phenotypic consequences. As such, HEs might have been actively selected for or against. Given that HEs formation is an ongoing process, one could imagine that HEs still constitute a source of genetic variation available for breeders. In Chalhoub et al. (2014), the authors reported the presence of HE in regions where QTLs have been found for traits such as oil biosynthesis, seed GSL content, disease resistance, and flowering that appear to be under breeding-directed selection.

Although the differential presence of HE between genotypes has been found to correlate with phenotypic variation in *B. napus*, the nature of the causal polymorphism (linked to gene expression or sequence) has rarely been investigated in details.

Chalhoub et al. (2014) proposed that difference in gene content between homoeologous A02/C02 regions may explain the presence of two quantitative trait loci (QTLs) for total aliphatic glucosinolates content. Given that the glucosinolate gene is absent from A02 (prior to the HE), an event ($A_n2^+ \rightarrow C_n2^-$) HE leads to the non-compensated loss of the gene on C02 involved in glucosinolate catabolism. A related example has been given by (Liu et al., 2012) who observed that an ($A_n9^+ \rightarrow C_n8^-$) HE resulted in duplication of a defective (A09) allele for the lignin biosynthesis gene *CCR1* and thus in reduction of the antinutritive fibre component in seed.

Less evidence has been found so far supporting a direct link between HE-driven differential gene expression and phenotypic consequences for traits of agronomic interest. Yet, differences in transcript abundance are indisputably a major contributor to phenotypic variation. In maize, 15% of the most extreme case of intraspecific differential expression (presence / absence of transcript – ePAVs) have been associated with agronomic and metabolic traits (Jin et al., 2016). Whereas we have clearly shown that large differences in expression are found within HE, we can expect that the presence of an HE could also impact gene expression outside of the HE (trans regulation). This is notably expected if transcriptions factors are involved in HE, which is currently the case. For example, (Schiessl et al., 2014) observed a copy number reduction affecting two FLC (FLOWERING LOCUS C) paralogs on chromosome C09, which is mirrored by a corresponding copy number increase on A10. FLC encodes a MADS box transcription factor that blocks flowering quantitatively by repressing the transcription of downstream floral pathway. As a consequence of this HE, longer vernalization is required to induce flowering. This example is not unique and in Chalhoub et al. (2014) the authors reported the presence of other FLC homologs affected by HEs. Thus the change in expression pattern for a single regulator gene within HE could possibly result in differential expression of multiple genes outside an HE.

6.3 The anti-CO activity of FANCM is conserved in the *Brassica*

The results obtained in both *B. rapa* and *B. napus* show that the anti CO activity of FANCM, first discovered in *A. thaliana*, is conserved in two representative crop species of the *Brassica*. We used a straightforward TILLING approach to identify mutations in the single copy of FANCM in *B. rapa* (*BraA.FANCM*) and in the two copies of FANCM present on either the A or the C subgenome of *B. napus (BnaA.FANCM* and *BnaC.FANCM,* respectively). A fair amount of mutations was detected in all three genes, ranging from 19 to 47; however, we did not find mutations inducing a stop codon in *BnaC.FANCM* (contrary to *BnaA.FANCM*; *BnaA.fancm-1*) and I could only use two mutations affecting strongly conserved amino acid substitutions (*BnaC.fancm-1* and *BnaC.fancm-2*). I also used a missense mutation in *B. rapa* (*Bra.fancm-1*) whereas two additional mutations leading to as stop codon were subsequently identified. Ongoing work aims at replicating the results obtained with *Bra.fancm-1*with these two harmful alleles.

In *B. rapa*, we showed that *Bra.fancm-1* was able to restore bivalent formation in a CO-defective mutant **(***braA.msh4-1*$^{-/-}$); we observed an almost two-fold increase in the number of bivalents and chiasmata in the double mutant (*braA.msh4-1*$^{-/-}$ *braA.fancm-1*$^{-/-}$) compared to the single mutant *braA.msh4-1*$^{-/-}$. This demonstrated that *BraA.FANCM* limits CO frequencies in *B. rapa.*

In *B. napus*, a modest (~30%) but consistent increase in CO frequencies was observed in one of the two mutants I analysed (*bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$). We showed that this slight increase affected both CO formation between homologous (in AACC; consistent across two independent intervals) and homoeologous (in AC) chromosomes. As discussed p.184, I have doubts as to whether the extra-COs observed in the AC plants are formed between homoeologous regions; we instead hypothesize that they are formed between homologous regions duplicated on homoeologous chromosomes as a consequence of HEs. This notwithstanding, these results suggest that FANCM has an antiCO effect in *B. napus*. Given that *bnaC.fancm-2* is likely not a complete loss of function mutation, the extent to which CO frequencies can be increased in *B. napus* remains an open question.

Although our results lend strong support to the anti-CO activity of FANCM in *Brassica*, I was a bit disappointed with the outcome of the analysis in *Brassica napus*. There is clearly experience to be learned from this translational biology approach and I think this general discussion is the appropriate place to discuss how the experimental design that we proposed can be improved.

First it appears necessary to comment on the major limiting factor of the analysis, i.e., the absence of a mutation inducing a stop in all copy of FANCM in *B. napus*. In retrospect, the odds against such an occurrence were ~25% given the size of the population and the length of the FANCM sequence that were screened for mutations. Our results thus call for an improved design that maximises the odds of finding mutations inducing STOP codon. High throughput approaches, like TILLING by sequencing (Tsai et al., 2011), are developing, including in *B. napus* (Gilchrist et al., 2013); these approaches provide means to multiply the odds of finding STOP codon. Using this approach, the screening of ~1000 M2 plants using a (summed) fragment of 2500bp should result in the detection of more than 6 stop mutations on average (the probability to detect no STOP being 0.001). Finding two stop mutations per copy (preferentially not in the last exons of the protein) would be a reasonable objective for this screen as this experimental design call for replicates.

The need for replicates in our assay was in part motivated by the high mutation load of the EMS-mutagenized population. The number of off-target mutations in our *fancm* mutant plants (~ 9000 that are predicted to affect protein function) question the risk of confusion between mutations in FANCM and off-target mutations (see discussion p.185). In *B. napus*, we assessed CO frequencies in plants that carry mutations for each of the two copies of *FANCM* (*BnaA.FANCM* and *BnaC.FANCM).* This first required combining individual (either A or C) mutant alleles into a single F1 plant and identifying double homozygote mutant plants in the selfed progeny of this F1 plant (i.e. F2 population). Producing these populations in *Brassica napus* has required a fair amount of time, effort and greenhouse space, thereby putting a practical limit to the number of mutation combinations that can be analysed. For example, choosing more than two mutations in each copy of *FANCM* would hardly have been feasible during my PhD. Thus, there must be as little ambiguity as possible on the outcome of these mutations on protein activity. In the absence of stop inducing mutations, Kumar et al. (2009) proposed a tool to carefully choose mutations that are predicted to affect the function of the protein. This is the tool I used to select *BraA.fancm*-1, *BnaC.fancm-1* and *BnaC.fancm-2*, which highlighted some of the limits of this approach.

My results illustrate how the use of missense mutations can introduce uncertainty in the issue, even when they target amino acid known to be important for the function of the protein. For example, *BnaC.fancm-2* proved to be a non-null mutation while it led to change a highly conserved amino acid in the helicase domain of FANCM, whose substitution is causal for a loss of function FANCM in *A. thaliana*. By contrast, although we were far less selective when it

came to choose missense mutations in *BraA.FANCM* (compared to *BnaC.FANCM*), it turned out that *BraA.fancm-1* resulted in a complete loss of function protein contrary to *BnaC.fancm-1* or *BnaC.fancm-2*. Surprisingly enough, although a total of 30 defective alleles of *FANCM* were recovered from the screening of ~7000 mutagenized lines of *A. thaliana* (R. Mercier, comm. pers.), none of these mutations have targeted the same amino-acid as in *BraA.fancm-1*. Altogether, these results indicate that the effect of missense mutations can hardly be generalized. Therefore, if I were to repeat this analysis, I would only focus on mutations inducing stop codons.

EMS-SNPs were instrumental to design our genetic assay, but introduced some (low) level of heterozygosity in the plants that we used for this assay. Given that the anti-CO activity of FANCM is broken down in hybrids (at least in *A. thaliana*; Girard et al. (2015), the presence of background EMS-SNPs raises question about their possible antagonistic effect on CO increase in *B. napus fancm* mutants (see discussion p.183) and has entirely undermined the work undertaken to evaluate the possible cumulative effect of *fancm* mutations in triploids (AAC) *Brassica* hybrids. Although this approach was abandoned in the course of my PhD work, the questions continues to be relevant but should be addressed using other anti-CO proteins.

Given what is known in *A. thaliana* about the pathways that limits CO frequencies, it would be worth repeating the complete experiment with genes involved in the two other anti-CO pathways. Evaluating the effect of mutations in *FIDL1* in triploids (AAC) *Brassica* hybrids is clearly relevant as FIDGL1 is thought to act upstream of all the other anti-CO proteins (Girard et al., 2015). In absolute terms, the most promising target for a new screen would be RECQ4A – RECQ4B whose concomitant depletion lead to a six-fold increase in CO frequency in *A. thaliana*. However, as we have seen before, RECQ4A and RECQ4B are present in four and two copies in *B. napus*, respectively. The choice of finding mutant in gene present in multiple copies though TILLING might seem debatable as CRISPR-CAS9 clearly surpass in theory this approach. This is clearly what need to be done but CRISPR-CAS9 in *Brassica* not yet available.

References

Adams, K.L., Cronn, R., Percifield, R., and Wendel, J.F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U. S. A. **100**: 4649–54.

Aggarwal, D.D., Rashkovetsky, E., Michalak, P., Cohen, I., Ronin, Y., Zhou, D., Haddad, G.G., and Korol, A.B. (2015). Experimental evolution of recombination and crossover interference in Drosophila caused by directional selection for stress-related traits. BMC Biol. **13**: 101.

Aklilu, B.B., Soderquist, R.S., and Culligan, K.M. (2014). Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. Nucleic Acids Res. **42**: 3104–3118.

Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S., and Moore, G. (2008). Detailed dissection of the chromosomal region containing the Ph1 locus in wheat Triticum aestivum: with deletion mutants and expression profiling. Ann. Bot. **101**: 863–72.

Albalat, R. and Cañestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. **17**: 379–391.

Albini, S.M. and Jones, G.H. (1987). Synaptonemal complex spreading in Allium cepa and A. fistulosum - I. The initiation and sequence of pairing. Chromosoma **95**: 324–338.

Allender, C.J. and King, G.J. (2010). Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol. **10**: 54.

Allers, T. and Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. Cell **106**: 47–57.

Alós, E., Roca, M., Iglesias, D.J., Mínguez-Mosquera, M.I., Damasceno, C.M.B., Thannhauser, T.W., Rose, J.K.C., Talón, M., and Cercós, M. (2008). An evaluation of the basis and consequences of a stay-green mutation in the navel negra citrus mutant using transcriptomic and proteomic profiling and metabolite analysis. Plant Physiol. **147**: 1300–1315.

Altman, D.W. and Busch, R.H. (1984). (1984) Random Intermating Before Selection in Spring Wheat. Components: 1085–1089.

Alves-Rodrigues, I., Ferreira, P.G., Moldón, A., Vivancos, A.P., Hidalgo, E., Guigó, R., and Ayté, J. (2016). Spatiotemporal Control of Forkhead Binding to DNA Regulates the

Meiotic Gene Expression Program. Cell Rep. **14**: 885–895.

Anderson, L.K., Lohmiller, L.D., Tang, X., Hammond, D.B., Javernick, L., Shearer, L., Basu-Roy, S., Martin, O.C., and Falque, M. (2014). Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers. PNAS Early Ed.

Andreuzza, S., Nishal, B., Singh, A., and Siddiqi, I. (2015). The Chromatin Protein DUET/MMD1 Controls Expression of the Meiotic Gene TDM1 during Male Meiosis in Arabidopsis. PLoS Genet. **11**: 1–22.

Armstrong, S.J., Caryl, A.P., Jones, G.H., and Franklin, F.C.H. (2002). Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. J. Cell Sci. **115**: 3645–3655.

Armstrong, S.J., Franklin, F.C.H., and Jones, G.H. (2003). A meiotic time-course for Arabidopsis thaliana. Sex. Plant Reprod. **16**: 141–149.

Arumuganathan, E.D. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. Plant Mol. Biol. Report. **9**: 208–218.

Auer, P.L. and Doerge, R.W. (2010). Statistical design and analysis of RNA sequencing data. Genetics **185**: 405–416.

Bai, Z. et al. (2016). The impact and origin of copy number variations in the Oryza species. BMC Genomics **17**: 261.

Barakate, A., Higgins, J.D., Vivera, S., Stephens, J., Perry, R.M., Ramsay, L., Colas, I., Oakey, H., Waugh, R., Franklin, F.C.H., Armstrong, S.J., and Halpin, C. (2014). The synaptonemal complex protein ZYP1 is required for imposition of meiotic crossovers in barley. Plant Cell **26**: 729–740.

Battagin, M., Gorjanc, G., Faux, A.-M., Johnston, S.E., and Hickey, J.M. (2016). Effect of manipulating recombination rates on response to selection in livestock breeding programs. Genet. Sel. Evol. **48**: 44.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science **327**: 836–840.

Bauer, E. et al. (2013). Intraspecific variation of recombination rate in maize. Genome Biol. **14**: R103.

Bertioli, D.J. et al. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat. Genet. **48**.

Bicknell, R.A. and Koltunow, A.M. (2004). Understanding apomixis: recent advances and

remaining conundrums. Plant Cell **16 Suppl**: S228-45.

Boden, S. a, Langridge, P., Spangenberg, G., and Able, J. a (2009). TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1. Plant J. **57**: 487–97.

Bomblies, K., Jones, G., Franklin, C., Zickler, D., and Kleckner, N. (2016). The challenge of evolving stable polyploidy: could an increase in crossover interference distance play a central role? Chromosoma **125**: 287–300.

Borde, V., Goldman, a S., and Lichten, M. (2000). Direct coupling between meiotic DNA replication and recombination initiation. Science **290**: 806–809.

Brown, M.S., Grubb, J., Zhang, A., Rust, M.J., and Bishop, D.K. (2015). Small Rad51 and Dmc1 Complexes Often Co-occupy Both Ends of a Meiotic DNA Double Strand Break. PLoS Genet. **11**.

Butruille, D. V et al. (1999). Linkage analysis of molecular markers and quantitative trait loci in populations of inbred backcross lines of Brassica napus L. Genetics **153**: 949–64.

Canady, M.A., Ji, Y., and Chetelat, R.T. (2006). Homeologous recombination in Solanum lycopersicoides introgression lines of cultivated tomato. Genetics **174**: 1775–1788.

Carlton, P.M., Farruggio, A.P., and Dernburg, A.F. (2006). A link between meiotic prophase progression and crossover control. PLoS Genet. **2**: 119–128.

Carvalho, C.M.B. and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. **17**: 224–238.

Chalhoub, B. et al. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science (80-.). **345**: 950–953.

Chelysheva, L.A., Grandont, L., and Grelon, M. (2013). Immunolocalization of meiotic proteins in Brassicaceae: method 1. Methods Mol. Biol. **990**: 93–101.

Chelysheva, L., Gendrot, G., Vezon, D., Doutriaux, M.P., Mercier, R., and Grelon, M. (2007). Zip4/Spo22 is required for class I CO formation but not for synapsis completion in Arabidopsis thaliana. PLoS Genet. **3**: 802–813.

Chelysheva, L., Grandont, L., Vrielynck, N., Le Guin, S., Mercier, R., and Grelon, M. (2010). An easy protocol for studying chromatin and recombination protein dynamics during Arabidopsisthaliana meiosis: Immunodetection of cohesins, histones and MLH1. Cytogenet. Genome Res. **129**: 143–153.

Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J., Smith, A.G., and Retzel, E.F. (2010). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. **10**: 280.

Cheng, F. et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in Brassica rapa and Brassica oleracea. Nat. Genet.: 1–10.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., and Wang, X. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of Brassica rapa. PLoS One **7**.

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of Brassica plants. Hortic. Res. **1**: 14024.

Chester, M., Gallagher, J.P., Symonds, V. V., Cruz da Silva, a. V., Mavrodiev, E. V., Leitch, a. R., Soltis, P.S., and Soltis, D.E. (2012a). Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proc. Natl. Acad. Sci. **109**: 1176–1181.

Chester, M., Gallagher, J.P., Symonds, V.V., Cruz da Silva, A.V., Mavrodiev, E. V, Leitch, A.R., Soltis, P.S., and Soltis, D.E. (2012b). Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proc. Natl. Acad. Sci. U. S. A. **109**: 1176–81.

Cheung, F., Trick, M., Drou, N., Lim, Y.P., Park, J.-Y., Kwon, S.-J., Kim, J.-A., Scott, R., Pires, J.C., Paterson, A.H., Town, C., and Bancroft, I. (2009). Comparative analysis between homoeologous genome segments of Brassica napus and its progenitor species reveals extensive sequence-level divergence. Plant Cell **21**: 1912–28.

Chia, J.M. et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. **44**: 803–807.

Choi, K., Zhao, X., Kelly, K.A., Venn, O., Higgins, J.D., Yelina, N.E., Hardcastle, T.J., Ziolkowski, P.A., Copenhaver, G.P., Franklin, F.C.H., McVean, G., and Henderson, I.R. (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. **45**: 1327–36.

Choulet, F. et al. (2014). Structural and functional partitioning of bread wheat chromosome 3B. Science **345**: 1249721.

Chowdhury, R., Bois, P.R.J., Feingold, E., Sherman, S.L., and Cheung, V.G. (2009). Genetic analysis of variation in human meiotic recombination. PLoS Genet. **5**.

Cifuentes, M., Eber, F., Lucas, M.-O., Lode, M., Chèvre, A.-M., and Jenczewski, E. (2010). Repeated polyploidy drove different levels of crossover suppression between homoeologous chromosomes in Brassica napus allohaploids. Plant Cell **22**: 2265–2276.

Clough, S.J. and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-

mediated transformation of Arabidopsis thaliana. Plant J. **16**: 735–743.

Coopera, T.J., Garcia, V., and Neale, M.J. (2016). Meiotic DSB patterning: A multifaceted process. Cell Cycle **15**: 13–21.

Crismani, W., Baumann, U., Sutton, T., Shirley, N., Webster, T., Spangenberg, G., Langridge, P., and Able, J. a (2006). Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. BMC Genomics **7**: 267.

Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J.L., Chelysheva, L., Copenhaver, G.P., Horlow, C., and Mercier, R. (2012). FANCM limits meiotic crossovers. Science **336**: 1588–90.

Crismani, W., Girard, C., and Mercier, R. (2013). Tinkering with meiosis. J. Exp. Bot. **64**: 55–65.

D'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., and Mercier, R. (2009). Turning meiosis into mitosis. PLoS Biol. **7**: e1000124.

Delourme, R. et al. (2013). High-density SNP-based genetic map development and linkage disequilibrium assessment in Brassica napus L. BMC Genomics **14**: 120.

Desai, A., Chee, P.W., Rong, J., May, O.L., and Paterson, A.H. (2006). Chromosome structural changes in diploid and tetraploid A genomes of Gossypium. Genome **49**: 336–345.

Dirks, R. et al. (2009). Reverse breeding: A novel breeding approach based on engineered meiosis. Plant Biotechnol. J. **7**: 837–845.

Dole, J. and Weber, D.F. (2007). Detection of quantitative trait loci influencing recombination using recombinant inbred lines. Genetics **177**: 2309–2319.

Dukowic-Schulze, S. and Chen, C. (2014). The meiotic transcriptome architecture of plants. Front. Plant Sci. **5**: 220.

Dukowic-Schulze, S., Harris, A., Li, J., Sundararajan, A., Mudge, J., Retzel, E.F., Pawlowski, W.P., and Chen, C. (2014a). Comparative Transcriptomics of Early Meiosis in Arabidopsis and Maize. J. Genet. Genomics **41**: 139–152.

Dukowic-Schulze, S., Sundararajan, A., Mudge, J., Ramaraj, T., Farmer, A.D., Wang, M., Sun, Q., Pillardy, J., Kianian, S., Retzel, E.F., Pawlowski, W.P., and Chen, C. (2014b). The transcriptome landscape of early maize meiosis. BMC Plant Biol. **14**: 118.

Dukowic-Schulze, S., Sundararajan, A., Ramaraj, T., Kianian, S., Pawlowski, W.P., Mudge, J., and Chen, C. (2016). Novel Meiotic miRNAs and Indications for a Role of PhasiRNAs in Meiosis. Front. Plant Sci. **7**: 762.

Ederveen, A., Lai, Y., van Driel, M.A., Gerats, T., and Peters, J.L. (2015). Modulating

crossover positioning by introducing large structural changes in chromosomes. BMC Genomics **16**: 89.

Esch, E., Szymaniak, J.M., Yates, H., Pawlowski, W.P., and Buckler, E.S. (2007). Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. Genetics **177**: 1851–1858.

Feuillet, C., Penger, A., Gellner, K., Mast, A., and Keller, B. (2001). Molecular Evolution of Receptor-Like Kinase Genes in Hexaploid Wheat. Independent Evolution of Orthologs after Polyploidization and Mechanisms of Local Rearrangements at Paralogous Loci. PLANT Physiol. **125**: 1304–1313.

Flórez-Zapata, N.M. V, Reyes-Valdés, M.H., Hernandez-Godínez, F., and Martínez, O. (2014). Transcriptomic landscape of prophase I sunflower male meiocytes. Front. Plant Sci. **5**: 277.

Flórez-Zapata, N.M.V., Reyes-Valdés, M.H., and Martínez, O. (2016). Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates. BMC Genomics: 1–16.

Flowers, J.M., Molina, J., Rubinstein, S., Huang, P., Schaal, B.A., and Purugganan, M.D. (2012). Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. Mol. Biol. Evol. **29**: 675–687.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu. Rev. Plant Biol. **60**: 433–453.

Gaeta, R.T. and Chris Pires, J. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol. **186**: 18–28.

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E., and Osborn, T.C. (2007). Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype. Plant Cell **19**: 3403–3417.

Gaeta, R.T., Yoo, S.-Y., Pires, J.C., Doerge, R.W., Chen, Z.J., and Osborn, T.C. (2009). Analysis of gene expression in resynthesized Brassica napus Allopolyploids using arabidopsis 70mer oligo microarrays. PLoS One **4**: e4760.

Garcia, V., Gray, S., Allison, R.M., Cooper, T.J., and Neale, M.J. (2015). Tel1(ATM)-mediated interference suppresses clustered meiotic double-strand-break formation. Nature **520**: 114–118.

Gari, K., Décaillet, C., Stasiak, A.Z., Stasiak, A., and Constantinou, A. (2008). The Fanconi Anemia Protein FANCM Can Promote Branch Migration of Holliday Junctions

and Replication Forks. Mol. Cell **29**: 141–148.

Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J., and Anderson, L.K. (2007). Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. **8**: 77–84.

Gazave, E., Tassone, E.E., Ilut, D.C., Wingerson, M., Datema, E., Witsenboer, H.M.A., Davis, J.B., Grant, D., Dyer, J.M., Jenks, M.A., Brown, J., and Gore, M.A. (2016). Population Genomic Analysis Reveals Differential Evolutionary Histories and Patterns of Diversity across Subgenomes and Subpopulations of Brassica napus L. Front. Plant Sci. **7**: 1–16.

Gilchrist, E.J., Sidebottom, C.H.D., Koh, C.S., Macinnes, T., Sharpe, A.G., and Haughn, G.W. (2013). A mutant Brassica napus (canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. PLoS One **8**: e84303.

Girard, C., Chelysheva, L., Choinard, S., Froger, N., Macaisne, N., Lehmemdi, A., Mazel, J., Crismani, W., and Mercier, R. (2015). AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. PLOS Genet. **11**: e1005369.

Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C., and Mercier, R. (2014). FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. Nucleic Acids Res.: 1–9.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O.C., and Mézard, C. (2011). Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. PLoS Genet. **7**: e1002354.

Glémin, S. (2010). Surprising fitness consequences of GC-biased gene conversion: I. Mutation load and inbreeding depression. Genetics **185**: 939–959.

Glover, N.M., Redestig, H., and Dessimoz, C. (2016). Homoeologs: What Are They and How Do We Infer Them? Trends Plant Sci. **xx**: 1–13.

Goddard, M.R., Charles, H., Godfray, J., and Burt, A. (2005). Sex increases the efficacy of natural selection in experimental yeast populations. Nature **434**: 636–640.

Gong, L., Olson, M., and Wendel, J.F. (2014). Cytonuclear evolution of rubisco in four allopolyploid lineages. Mol. Biol. Evol. **31**: 2624–36.

Grandont, L., Cuñado, N., Coriton, O., Huteau, V., Eber, F., Chèvre, A.M., Grelon, M., Chelysheva, L., and Jenczewski, E. (2014). Homoeologous Chromosome Sorting and Progression of Meiotic Recombination in Brassica napus: Ploidy Does Matter! Plant Cell **26**: 1448–1463.

Greer, E., Martin, A.C., Pendle, A., Colas, I., Jones, A.M.E., Moore, G., and Shaw, P. (2012). The Ph1 Locus Suppresses Cdk2-Type Activity during Premeiosis and Meiosis in Wheat. Plant Cell **24**: 152–162.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I., and Moore, G. (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature **439**: 749–752.

Guo, H., Wang, X., Gundlach, H., Mayer, K.F.X., Peterson, D.G., Scheffler, B.E., Chee, P.W., and Paterson, A.H. (2014). Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, Gossypium (Cotton). Genetics **197**: 1153–1163.

Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S. a a C., Cook, S., Pravenec, M., Aitman, T., Jacob, H., Shull, J.D., Hubner, N., and Cuppen, E. (2008). Distribution and functional impact of DNA copy number variation in the rat. Nat. Genet. **40**: 538–45.

Hamant, O., Ma, H., and Cande, W.Z. (2006). Genetics of meiotic prophase I in plants. Annu. Rev. Plant Biol. **57**: 267–302.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., and Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species Brassica napus. Nat. Biotechnol. **30**: 798–802.

Hasan, M., Seyis, F., Badani, A.G., Pons-Kühnemann, J., Friedt, W., Lühs, W., and Snowdon, R.J. (2006). Analysis of genetic diversity in the Brassica napus L. gene pool using SSR markers. Genet. Resour. Crop Evol. **53**: 793–802.

Hayashi, M., Mlynarczyk-Evans, S., and Villeneuve, A.M. (2010). The synaptonemal complex shapes the crossover landscape through cooperative assembly, crossover promotion and crossover inhibition during Caenorhabditis elegans meiosis. Genetics **186**: 45–58.

He, Z., Cheng, F., Li, Y., Wang, X., Parkin, I. a. P., Chalhoub, B., Liu, S., and Bancroft, I. (2015). Construction of Brassica A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. Data Br. **4**: 357–362.

Henrichsen, C.N., Chaignat, E., and Reymond, A. (2009). Copy number variants, diseases and gene expression. Hum. Mol. Genet. **18**: R1-8.

Henry, I.M., Dilkes, B.P., Tyagi, A., Gao, J., Christensen, B., and Comai, L. (2014). The BOY NAMED SUE Quantitative Trait Locus Confers Increased Meiotic Stability to an Adapted Natural Allopolyploid of Arabidopsis. Plant Cell Online **26**: 181–194.

Higgins, J., Magusin, A., Trick, M., Fraser, F., and Bancroft, I. (2012a). Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. BMC Genomics **13**: 247.

Higgins, J.D., Armstrong, S.J., Franklin, F.C.H., and Jones, G.H. (2004). The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: Evidence for two classes of recombination in Arabidopsis. Genes Dev. **18**: 2557–2570.

Higgins, J.D., Buckling, E.F., Franklin, F.C.H., and Jones, G.H. (2008). Expression and functional analysis of AtMUS81 in Arabidopsis meiosis reveals a role in the second pathway of crossing-over. Plant J. **54**: 152–162.

Higgins, J.D., Osman, K., Jones, G.H., and Franklin, F.C.H. (2014). Factors Underlying Restricted Crossover Localization in Barley Meiosis. Annu. Rev. Genet.: 29–47.

Higgins, J.D., Perry, R.M., Barakate, A., Ramsay, L., Waugh, R., Halpin, C., Armstrong, S.J., and Franklin, F.C.H. (2012b). Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley. Plant Cell **24**: 4096–4109.

Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., Jones, G.H., and Franklin, F.C.H. (2005). The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. Genes Dev. **19**: 2488–500.

Hobolth, P. (1981). Chromosome pairing in allohexaploid wheat var. Chinese Spring. Transformation of multivalents into bivalents, a mechanism for exclusive bivalent formation. Carlsberg Res. Commun. **46**: 129–173.

Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A Time-Calibrated Road Map of Brassicaceae Species Radiation and Evolutionary History. Plant Cell **27**: 2770–84.

Hose, J., Yong, C.M., Sardi, M., Wang, Z., Newton, M.A., and Gasch, A.P. (2015). Dosage compensation can buffer copy-number variation in wild yeast. Elife **4**.

Howell, E.C., Kearsey, M.J., Jones, G.H., King, G.J., and Armstrong, S.J. (2008). A and C genome distinction and chromosome identification in brassica napus by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics **180**: 1849–57.

Hunter, N. (2015). Meiotic recombination: The essence of heredity. Cold Spring Harb. Perspect. Biol. **7**.

International Barley Genome Sequencing, C. et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. Nature **491**: 711–716.

International Wheat Genome Sequencing Consortium (IWGSC), T.I.W.G.S.C. et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science **345**: 1251788.

Jahns, M.T., Vezon, D., Chambon, A., Pereira, L., Falque, M., Martin, O.C., Chelysheva, L., and Grelon, M. (2014). Crossover localisation is regulated by the neddylation posttranslational regulatory pathway. PLoS Biol. **12**: e1001930.

Jannoo, N., Grivet, L., David, J., D'Hont, A., and Glaszmann, J.-C. (2004). Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. Heredity (Edinb). **93**: 460–467.

Jenczewski, E. (2013). Evolution: He who grabs too much loses all. Curr. Biol. **23**.

Jenczewski, E. and Alix, K. (2004). From Diploids to Allopolyploids: The Emergence of Efficient Pairing Control Genes in Plants. CRC. Crit. Rev. Plant Sci. **23**: 21–45.

Jenczewski, E., Eber, F., Grimaud, A., Huet, S., Lucas, M.O., Monod, H., and Chèvre, A.M. (2003). PrBn, a major gene controlling homeologous pairing in oilseed rape (Brassica napus) haploids. Genetics **164**: 645–653.

Jenkins, G., and White, J. (1988). Elimination of multivalents during meiotic prophase in Scilla autumnalis. II. Tetraploid. Genome 30: 940–946.

Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., and Wang, Y. (2016). Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. Nat. Publ. Gr.: 1–12.

Jones, L.E., Rybka, K., and Lukaszewski, A.J. (2002). The effect of a deficiency and a deletion on recombination in chromosome 1BL in wheat. Theor. Appl. Genet. **104**: 1204–1208.

Joyce, E.F., Pedersen, M., Tiong, S., White-Brown, S.K., Paul, A., Campbell, S.D., and McKim, K.S. (2011). Drosophila ATM and ATR have distinct activities in the regulation of meiotic DNA damage and repair. J. Cell Biol. **195**: 359–367.

Kauppi, L., Barchi, M., Lange, J., Baudat, F., Jasin, M., and Keeney, S. (2013). Numerical constraints and feedback control of double-strand breaks in mouse meiosis. Genes Dev. **27**: 873–886.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell **88**: 375–384.

Keeney, S., Lange, J., and Mohibullah, N. (2014). Self-organization of meiotic recombination initiation: general principles and molecular pathways. Annu. Rev. Genet.

48: 187–214.

Kessner, D. and Novembre, J. (2015). Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. Genetics **199**: 991–1005.

Kleckner, N., Zickler, D., Jones, G.H., Dekker, J., Padmore, R., Henle, J., and Hutchinson, J. (2004). A mechanical basis for chromosome function. Proc. Natl. Acad. Sci. U. S. A. **101**: 12592–7.

Knight, E., Greer, E., Draeger, T., Thole, V., Reader, S., Shaw, P., and Moore, G. (2010). Inducing chromosome pairing through premature condensation: Analysis of wheat interspecific hybrids. Funct. Integr. Genomics **10**: 603–608.

Knoll, A., Higgins, J.D., Seeliger, K., Reha, S.J., Dangel, N.J., Bauknecht, M., Schröpfer, S., Franklin, F.C.H., and Puchta, H. (2012). The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis. Plant Cell **24**: 1448–64.

Knoll, A. and Puchta, H. (2011). The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. J. Exp. Bot. **62**: 1565–1579.

Koh, J., Soltis, P.S., and Soltis, D.E. (2010). Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid Tragopogon mirus (Asteraceae). BMC Genomics **11**: 97.

Koltunow, A.M.G. et al. (2011). Sexual reproduction is the default mode in apomictic Hieracium subgenus Pilosella, in which two dominant loci function to enable apomixis. Plant J. **66**: 890–902.

Kong, A. et al. (2008). Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science **319**: 1398–1401.

Kong, A., Thorleifsson, G., Frigge, M.L., Masson, G., Gudbjartsson, D.F., Villemoes, R., Magnusdottir, E., Olafsdottir, S.B., Thorsteinsdottir, U., and Stefansson, K. (2014). Common and low-frequency variants associated with genome-wide recombination rate. Nat. Genet. **46**: 11–6.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. **4**: 1073–1081.

Lam, I. and Keeney, S. (2015). Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science (80-.). **350**: 932–937.

Lange, J., Pan, J., Cole, F., Thelen, M.P., Jasin, M., and Keeney, S. (2011). ATM controls meiotic double-strand-break formation. Nature **479**: 237–40.

Lashermes, P., Combes, M.C., Hueber, Y., Severac, D., and Dereeper, A. (2014). Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. Plant J. **78**: 674–685.

Lashermes, P., Hueber, Y., Combes, M.-C., Severac, D., and Dereeper, A. (2016). Inter-genomic DNA Exchanges and Homoeologous Gene Silencing Shaped the Nascent Allopolyploid Coffee Genome (Coffea arabica L.). G3: Genes|Genomes|Genetics **6**: 2937–2948.

Leflon, M., Grandont, L., Eber, F., Huteau, V., Coriton, O., Chelysheva, L., Jenczewski, E., and Chèvre, A.-M. (2010). Crossovers get a boost in Brassica allotriploid and allotetraploid hybrids. Plant Cell **22**: 2253–64.

Lenormand, T. and Dutheil, J. (2005). Recombination difference between sexes: A role for haploid selection. In PLoS Biology, pp. 0396–0403.

Li, F. et al. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat. Biotechnol. **33**.

Li, J., Farmer, A.D., Lindquist, I.E., Dukowic-Schulze, S., Mudge, J., Li, T., Retzel, E.F., and Chen, C. (2012). Characterization of a set of novel meiotically-active promoters in Arabidopsis. BMC Plant Biol. **12**: 104.

Li, X., Chang, Y., Xin, X., Zhu, C., Li, X., Higgins, J.D., and Wu, C. (2013). Replication protein A2c coupled with replication protein A1c regulates crossover formation during meiosis in rice. Plant Cell **25**: 3885–99.

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., and De Smet, R. (2016). Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. Plant Cell **28**: 326–344.

Liharska, T., Wordragen, M., Kammen, a, Zabel, P., and Koornneef, M. (1996). Tomato chromosome 6: effect of alien chromosomal segments on recombinant frequencies. Genome **39**: 485–91.

Lim, K.Y., Soltis, D.E., Soltis, P.S., Tate, J., Matyasek, R., Srubarova, H., Kovarik, A., Pires, J.C., Xiong, Z., and Leitch, A.R. (2008). Rapid chromosome evolution in recently formed polyploids in Tragopogon (Asteraceae). PLoS One **3**: e3353.

Liu, L., Stein, A., Wittkop, B., Sarvari, P., Li, J., Yan, X., Dreyer, F., Frauen, M., Friedt, W., and Snowdon, R.J. (2012). A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in Brassica napus seeds. Theor. Appl. Genet. **124**: 1573–1586.

Liu, S. et al. (2014). The Brassica oleracea genome reveals the asymmetrical evolution of

polyploid genomes. Nat. Commun. **5**: 3930.

Liu, Z., Adamczyk, K., Manzanares-Dauleux, M., Eber, F., Lucas, M.-O., Delourme, R., Chèvre, A.M., and Jenczewski, E. (2006). Mapping PrBn and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (Brassica napus L.) haploids. Genetics **174**: 1583–96.

Lloyd, A.H. (2014). Meiotic gene evolution: can you teach a new dog new tricks? Mol Biol Evol.

Lloyd, A.H. et al. (2014). Meiotic gene evolution: Can you teach a new dog new tricks? Mol. Biol. Evol. **31**: 172Lloyd, A.H. et (2014). Meiotic gene evoluti.

Lombard, V. and Delourme, R. (2001). A consensus linkage map for rapeseed (Brassica napus L.): construction and integration of three individual maps from DH populations. TAG Theor. Appl. Genet. **103**: 491–507.

López, E., Pradillo, M., Oliver, C., Romero, C., Cuñado, N., and Santos, J.L. (2012). Looking for natural variation in chiasma frequency in Arabidopsis thaliana. J. Exp. Bot. **63**: 887–894.

Lorieux, M. (2012). MapDisto: Fast and efficient computation of genetic linkage maps. Mol. Breed. **30**: 1231–1235.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.

Lu, F. et al. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. Nat. Commun. **6**: 6914.

Lukaszewski, A.J. (2000). Manipulation of the 1RS.1BL translocation in wheat by induced homoeologous recombination. Crop Sci. **40**: 216–225.

Lukaszewski, A.J. (2008). Unexpected behavior of an inverted rye chromosome arm in wheat. Chromosoma **117**: 569–578.

Lukaszewski, A.J. and Kopecký, D. (2010). The Ph1 locus from wheat controls meiotic chromosome pairing in autotetraploid rye (Secale cereale L.). Cytogenet. Genome Res. **129**: 117–123.

Lukaszewski, A.J., Kopecky, D., and Linc, G. (2012). Inversions of chromosome arms 4AL and 2BS in wheat invert the patterns of chiasma distribution. Chromosoma **121**: 201–208.

Lysak, M.A., Koch, M.A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. Genome Res. **15**: 516–525.

Ma, L., O'Connell, J.R., VanRaden, P.M., Shen, B., Padhi, A., Sun, C., Bickhart, D.M.,

Cole, J.B., Null, D.J., Liu, G.E., Da, Y., and Wiggans, G.R. (2015). Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet. **11**.

Macaisne, N. (2010) Analyse de gènes impliqués dans la formation des crossing-over méiotiques chez la plante modèle Arabidopsis thaliana, thèse de doctorat en Sciences biologiques, Sciences du végétal. Soutenue à l'Université de Paris-Sud, faculté des Sciences d'Orsay.

MacQueen, A.J. (2015). Catching a (Double-Strand) Break: The Rad51 and Dmc1 Strand Exchange Proteins Can Co-occupy Both Ends of a Meiotic DNA Double-Strand Break. PLoS Genet. **11**: 1–4.

Malkova, A. and Haber, J.E. (2012). Mutations arising during repair of chromosome breaks. Annu. Rev. Genet. **46**: 455–73.

Marimuthu, M.P.A. et al. (2011). Synthetic clonal reproduction through seeds. Science **331**: 876.

Martín, A.C., Shaw, P., Phillips, D., Reader, S., and Moore, G. (2014). Licensing MLH1 sites for crossover during meiosis. Nat. Commun. **5**: 4580.

Martinez, M., Cuñado, N., Carcelén, N., and Romero, C. (2001). The Ph1 and Ph2 loci play different roles in the synaptic behaviour of hexaploid wheat Triticum aestivum. Theor. Appl. Genet. **103**: 398–405.

Mason, A.S., Rousseau-gueutin, M., Morice, J., and Bayer, P.E. (2015). Centromere Locations in Brassica A and C Genomes Revealed Through Half- Tetrad Analysis. **202**: 1–31.

Mata, J., Wilbrey, A., and Bähler, J. (2007). Transcriptional regulatory network for sexual differentiation in fission yeast. Genome Biol. **8**: R217.

McClosky, B. and Tanksley, S.D. (2013). The impact of recombination on short-term selection gain in plant breeding experiments. Theor. Appl. Genet. **126**: 2299–2312.

McMullen, M.D. et al. (2009). Genetic properties of the maize nested association mapping population. Science **325**: 737–40.

Melchinger, A.E., Geiger, H.H., Utz, H.F., and Schnell, F.W. (2003). Effect of recombination in the parent populations on the means and combining ability variances in hybrid populations of maize (Zea mays L.). Theor. Appl. Genet. **106**: 332–340.

Mercier, R., Jolivet, S., Vezon, D., Huppe, E., Chelysheva, L., Giovanni, M., Nogué, F., Doutriaux, M.P., Horlow, C., Grelon, M., and Mézard, C. (2005). Two meiotic crossover classes cohabit in Arabidopsis: One is dependent on MER3, whereas the other

one is not. Curr. Biol. **15**: 692–701.

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., and Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. Annu. Rev. Plant Biol. **66**: 141210140145001–.

Mézard, C., Tagliaro Jahns, M., and Grelon, M. (2015). Where to cross? New insights into the location of meiotic crossovers. Trends Genet.: 1–9.

Min, X.J., Butler, G., Storms, R., and Tsang, A. (2005). OrfPredictor: Predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. **33**.

Ming, R. and Man Wai, C. (2015). Assembling allopolyploid genomes: no longer formidable. Genome Biol. **16**: 27.

Moose, S.P. and Mumm, R.H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. **147**: 969–977.

Munoz-Fuentes, V. et al. (2015). Strong artificial selection in domestic mammals did not result in an increased recombination rate. Mol. Biol. Evol. **32**: 510–523.

Muyle, A., Serres-Giardi, L., Ressayre, A., Escobar, J., and Glémin, S. (2011). GC-biased gene conversion and selection affect GC content in the oryza genus (rice). Mol. Biol. Evol. **28**: 2695–2706.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J. Biosci. Bioeng. **104**: 34–41.

Nasmyth, K. and Haering, C.H. (2009). Cohesin: Its Roles and Mechanisms. Annu. Rev. Genet. **43**: 525–558.

Nasmyth, K. and Haering, C.H. (2005). The structure and function of SMC and kleisin complexes. Annu Rev Biochem **74**: 595–648.

Neale, M.J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature **436**: 1053–7.

Nicolas, S.D. et al. (2007). Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of Brassica napus haploids. Genetics **175**: 487–503.

Nicolas, S.D., Leflon, M., Monod, H., Eber, F., Coriton, O., Huteau, V., Chèvre, A.-M., and Jenczewski, E. (2009). Genetic regulation of meiotic cross-overs between related genomes in Brassica napus haploids and hybrids. Plant Cell **21**: 373–85.

Nicolas, S.D., Monod, H., Eber, F., Chèvre, A.-M., and Jenczewski, E. (2012). Non-random distribution of extensive chromosome rearrangements in Brassica napus depends

on genome organization. Plant J. **70**: 691–703.

Nogué, F., Mara, K., Collonnier, C., and Casacuberta, J.M. (2016). Genome engineering and plant breeding: impact on trait discovery and development. Plant Cell Rep.: 1475–1486.

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. Genetics **154**: 923–929.

Nordborg, M. et al. (2005). The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol. **3**: 1289–1299.

Ohkura, H. (2015). Meiosis : An Overview of Key Differences from Mitosis.: 1–14.

Osborn, T.C., Butrulle, D. V, Sharpe, A.G., Pickering, K.J., Parkin, I. a P., Parker, J.S., and Lydiate, D.J. (2003). Detection and effects of a homeologous reciprocal transposition in Brassica napus. Genetics **165**: 1569–77.

Otto, S.P. and Barton, N.H. (1997). The evolution of recombination: Removing the limits to natural selection. Genetics **147**: 879–906.

Otto, S.P. and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat Rev Genet **3**: 252–261.

Page, J.T., Liechty, Z.S., Alexander, R.H., Clemons, K., Hulse-Kemp, A.M., Ashrafi, H., Van Deynze, A., Stelly, D.M., and Udall, J.A. (2016). DNA Sequence Evolution and Rare Homoeologous Conversion in Tetraploid Cotton. PLOS Genet. **12**: e1006012.

Page, S.L. and Hawley, R.S. (2003). Chromosome Choreography: The Meiotic Ballet. **301**: 785–790.

Pala, I. et al. (2008). Dosage compensation by gene-copy silencing in a triploid hybrid fish. Curr. Biol. **18**: 1344–8.

Parkin, I.A. et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol **15**: R77.

Peciña, A., Smith, K.N., Mézard, C., Murakami, H., Ohta, K., and Nicolas, A. (2002). Targeted stimulation of meiotic recombination. Cell **111**: 173–184.

Pecinka, A., Fang, W., Rehmsmeier, M., Levy, A. a, and Mittelsten Scheid, O. (2011). Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol. **9**: 24.

Perry, G.H., Tchinda, J., McGrath, S.D., Zhang, J., Picker, S.R., Cáceres, A.M., Iafrate, A.J., Tyler-Smith, C., Scherer, S.W., Eichler, E.E., Stone, A.C., and Lee, C. (2006). Hotspots for copy number variation in chimpanzees and humans. Proc. Natl. Acad. Sci. U. S. A. **103**: 8006–11.

Pezer, Ž., Harr, B., Teschke, M., Babiker, H., and Tautz, D. (2015). Divergence patterns of genic copy number variation in natural populations of the house mouse (Mus musculus domesticus) reveal three conserved genes with major population-specific expansions. Genome Res. **25**: 1114–24.

Pfeiffer, T.W. (1993). Recombination rates of soybean varieties from different periods of introduction and release. Theor. Appl. Genet. **86**: 557–561.

Piper, T.E. and Fehr, W.R. (1987). Yield Improvement in a Soybean Population by Utilizing Alternative Strategies of Recurrent Selection. Crop Sci. **27**: 172–178.

Piquemal, J., Cinquin, E., Couton, F., Rondeau, C., Seignoret, E., doucet, I., Perret, D., Villeger, M.-J., Vincourt, P., and Blanchard, P. (2005). Construction of an oilseed rape (Brassica napus L.) genetic map with SSR markers. Theor. Appl. Genet. **111**: 1514–1523.

Pires, J., Zhao, J., Schranz, M., Leon, E., Quijada, P., Lukens, L., and Osborn, T. (2004). Flowering time divergence and genomic rearrangements in resynthesized Brassicapolyploids (Brassicaceae). Biol. J. Linn. Soc. **82**: 675–688.

Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G. V, and Camerini-Otero, R.D. (2014). DNA recombination. Recombination initiation maps of individual human genomes. Science **346**: 1256442.

Prieler, S., Penkner, A., Borde, V., and Klein, F. (2005). The control of Spo11 ' s interaction with meiotic recombination hotspots. Genes Dev. **19**: 255–269.

Primig, M., Williams, R.M., Winzeler, E.A., Tevzadze, G.G., Conway, A.R., Hwang, S.Y., Davis, R.W., and Esposito, R.E. (2000). The core meiotic transcriptome in budding yeasts. Nat. Genet. **26**: 415–423.

Pupilli, F. and Barcaccia, G. (2012). Cloning plants by seeds: Inheritance models and candidate genes to increase fundamental knowledge for engineering apomixis in sexual crops. J. Biotechnol. **159**: 291–311.

Purugganan, M.D. and Fuller, D.Q. (2009). The nature of selection during plant domestication. Nature **457**: 843–848.

Qi, L.L., Friebe, B., and Gill, B.S. (2002). A strategy for enhancing recombination in proximal regions of chromosomes. Chromosom. Res. **10**: 645–654.

Qian, L., Qian, W., and Snowdon, R.J. (2014). Sub-genomic selection patterns as a signature of breeding in the allopolyploid Brassica napus genome. **15**: 1–17.

Ramsey, J. and Schemske, D.W. (2002). Neopolyploidy in Flowering Plants. Annu. Rev. Ecol. Syst. **33**: 589–639.

Rattray, A., Santoyo, G., Shafer, B., and Strathern, J.N. (2015). Elevated Mutation Rate during Meiosis in Saccharomyces cerevisiae. PLoS Genet. **11**: e1004910.

Ravi, M. and Chan, S.W.L. (2010). Haploid plants produced by centromere-mediated genome elimination. Nature **464**: 615–8.

Rey, M.-D., Calderón, M.C., and Prieto, P. (2015). The use of the ph1b mutant to induce recombination between the chromosomes of wheat and barley. Front. Plant Sci. **6**: 1–9.

Reynolds, A. et al. (2013). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat. Genet. **45**: 269–78.

Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J., Glaubitz, J.C., Acharya, C.B., Mitchell, S.E., Li, C., Li, Y., and Buckler, E.S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc. Natl. Acad. Sci. U. S. A. **112**: 3823–8.

Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M., and Saravanan, R.S. (2004). Tackling the plant proteome: Practical approaches, hurdles and experimental tools. Plant J. **39**: 715–733.

Roselius, K., Stephan, W., and Städler, T. (2005). The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics **171**: 753–763.

Ruzycki, P.A., Tran, N.M., Kefalov, V.J., Kolesnikov, A. V, and Chen, S. (2015). Graded gene expression changes determine phenotype severity in mouse models of CRX-associated retinopathies. Genome Biol. **16**: 171.

Sailer, C., Schmid, B., and Grossniklaus, U. (2016). Apomixis allows the transgenerational fixation of phenotypes in hybrid plants. Curr. Biol. **26**: 331–337.

Sanchez-Moran, E., Armstrong, S.J., Santos, J.L., Franklin, F.C.H., and Jones, G.H. (2002). Variation in chiasma frequency among eight accessions of Arabidopsis thaliana. Genetics **162**: 1415–1422.

Sanchez-Moran, E., Santos, J.L., Jones, G.H., and Franklin, F.C.H. (2007). ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. Genes Dev. **21**: 2220–2233.

Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. PLoS Genet. **8**: e1002854.

Sankoff, D., Zheng, C., and Zhu, Q. (2010). The collapse of gene complement following whole genome duplication. BMC Genomics **11**: 313.

Sato, S. et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641.

Saxena, R.K., Edwards, D., and Varshney, R.K. (2014). Structural variations in plant genomes. Brief. Funct. Genomics **13**: 296–307.

Schiessl, S., Samans, B., HÃ¼ttel, B., Reinhard, R., and Snowdon, R.J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species Brassica napus. Front. Plant Sci. **5**: 404.

Séguéla-Arnaud, M. et al. (2015). Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proc. Natl. Acad. Sci. U. S. A. **112**: 4713–8.

Sehrish, T., Symonds, V.V., Soltis, D.E., Soltis, P.S., and Tate, J.A. (2015). Cytonuclear Coordination Is Not Immediate upon Allopolyploid Formation in Tragopogon miscellus (Asteraceae) Allopolyploids. PLoS One **10**: e0144339.

Sharpe, a G., Parkin, I. a, Keith, D.J., and Lydiate, D.J. (1995). Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (Brassica napus). Genome **38**: 1112–21.

Shi, J. and Lai, J. (2015). Patterns of genomic changes with crop domestication and breeding. Curr. Opin. Plant Biol. **24**: 47–53.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat. Biotechnol. **23**: 75–81.

Smagulova, F., Gregoretti, I. V., Brick, K., Khil, P., Camerini-Otero, R.D., and Petukhova, G. V. (2011). Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature **472**: 375–8.

Springer, N.M. et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. **5**.

Stack, S.M. and Anderson, L.K. (2002). Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize. Chromosom. Res. **10**: 329–345.

Stefaniak, T.R., Hyten, D.L., Pantalone, V.R., Klarer, A., and Pfeiffer, T.W. (2006). Soybean cultivars resulted from more recombination events than unselected lines in the same population. Crop Sci. **46**: 43–51.

Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G.J., and Østergaard, L. (2010). A rich TILLING resource for studying gene function in Brassica rapa. BMC

Plant Biol.

Storlazzi, A., Tesse, S., Ruprich-Robert, G., Gargano, S., Pöggeler, S., Kleckner, N., and Zickler, D. (2008). Coupling meiotic chromosome axis integrity to recombination. Genes Dev. **22**: 796–809.

Stronghill, P.E., Azimi, W., and Hasenkampf, C. a (2014). A novel method to follow meiotic progression in Arabidopsis using confocal microscopy and 5-ethynyl-2'-deoxyuridine labeling. Plant Methods **10**: 33.

Suay, L. et al. (2014). Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in Brassica. New Phytol.

Suay, L. et al. (2013). Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in Brassica. New Phytol.: 645–656.

Sutton, T., Whitford, R., Baumann, U., Dong, C., Able, J.A., and Langridge, P. (2003). The Ph2 pairing homoeologous locus of wheat (Triticum aestivum): Identification of candidate meiotic genes using a comparative genetics approach. Plant J. **36**: 443–456.

Suwabe, K., Morgan, C., and Bancroft, I. (2008). Integration of Brassica A genome genetic linkage map between Brassica napus and B. rapa. Genome **51**: 169–176.

Symington, L.S. and Heyer, W.D. (2006). Some disassembly required: Role of DNA translocases in the disruption of recombination intermediates and dead-end complexes. Genes Dev. **20**: 2479–2486.

Szadkowski, E. et al. (2010). The first meiosis of resynthesized Brassica napus, a genome blender. New Phytol. **186**: 102–112.

Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Coriton, O., Jenczewski, E., and Chèvre, A.M. (2011). Polyploid formation pathways have an impact on genetic rearrangements in resynthesized Brassica napus. New Phytol. **191**: 884–894.

Tate, J.A., Ni, Z., Scheen, A.C., Koh, J., Gilbert, C.A., Lefkowitz, D., Chen, Z.J., Soltis, P.S., and Soltis, D.E. (2006). Evolution and expression of homeologous loci in Tragopogon miscellus (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics **173**: 1599–1611.

Tenaillon, M.I. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc. Natl Acad. Sci. USA **98**: 9161–9166.

Tenaillon, M.I., Sawkins, M.C., Anderson, L.K., Stack, S.M., Doebley, J., and Gaut, B.S. (2002). Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp. mays L.). Genetics **162**: 1401–1413.

Tenaillon, M.I., U'Ren, J., Tenaillon, O., and Gaut, B.S. (2004). Selection versus demography: A multilocus investigation of the domestication process in maize. Mol. Biol. Evol. **21**: 1214–1225.

Tiley, G.P. and Burleigh, J.G. (2015). The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol Biol **15**: 194.

Timmermans, M.C.P., Das, O.P., Bradeen, J.M., and Messing, J. (1997). Region-specific cis- and trans-acting factors contribute to genetic variability in meiotic recombination in maize. Genetics **146**: 1101–1113.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. **7**: 562–78.

Trick, M., Long, Y., Meng, J., and Bancroft, I. (2009). Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnol. J. **7**: 334–46.

Tsai, H. et al. (2011). Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol. **156**: 1257–68.

Udall, J. a, Quijada, P. a, and Osborn, T.C. (2005). Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of Brassica napus L. Genetics **169**: 967–79.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mézard, C., Mayer, C., and Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. Science **351**: 939–43.

Wang, A.T. and Smogorzewska, A. (2015). SnapShot: Fanconi anemia and associated proteins. Cell **160**: 354–354.e1.

Wang, J., Niu, B., Huang, J., Wang, H., Yang, X., Dong, A., Makaroff, C.A., Ma, H., and Wang, Y. (2016a). The PHD Finger Protein MMD1/DUET Ensures the Progression of Male Meiotic Chromosome Condensation and Directly Regulates the Expression of the Condensin Gene CAP-D3. Plant Cell: tpc.00040.2016.

Wang, J., Street, N.R., Scofield, D.G., and Ingvarsson, P.K. (2016b). Natural Selection and Recombination Rate Variation Shape Nucleotide Polymorphism Across the Genomes of Three Related Populus Species. Genetics **202**: genetics.115.183152-.

Wang, J., Tian, L., Lee, H.-S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn,

T.C., Doerge, R.W., Comai, L., and Chen, Z.J. (2006). Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics **172**: 507–17.

Wang, K., Tang, D., Wang, M., Lu, J., Yu, H., Liu, J., Qian, B., Gong, Z., Wang, X., Chen, J., Gu, M., and Cheng, Z. (2009). MER3 is required for normal meiotic crossover formation, but not for presynaptic alignment in rice. J. Cell Sci. **122**: 2055–63.

Wang, K., Wang, M., Tang, D., Shen, Y., Qin, B., Li, M., and Cheng, Z. (2010a). PAIR3, an axis-associated protein, is essential for the recruitment of recombination elements onto meiotic chromosomes in rice. Mol. Biol. Cell **22**: 12–19.

Wang, M., Wang, K., Tang, D., Wei, C., Li, M., Shen, Y., Chi, Z., Gu, M., and Cheng, Z. (2010b). The central element protein ZEP1 of the synaptonemal complex regulates the number of crossovers during meiosis in rice. Plant Cell **22**: 417–30.

Wang, N., Wang, Y., Tian, F., King, G.J., Zhang, C., Long, Y., Shi, L., and Meng, J. (2008). A functional genomics resource for\n Brassica napus\n : development of an EMS mutagenized population and discovery of\n FAE1\n point mutations by TILLING. New Phytol. **180**: 751–765.

Wang, S., Zickler, D., Kleckner, N., and Zhang, L. (2015). Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process. Cell Cycle **14**: 305–314.

Wang, X. et al. (2011). The genome of the mesopolyploid crop species Brassica rapa. Nat. Genet. **43**: 1035–1039.

Webster, M.T. and Hurst, L.D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. Trends Genet. **28**: 101–109.

Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013a). Phenotypic impact of genomic structural variation: insights from and for human disease. Nat. Rev. Genet. **14**: 125–138.

Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013b). Phenotypic impact of genomic structural variation: insights from and for human disease. Nat. Rev. Genet. **14**: 125–38.

Wendel, J.F., Jackson, S.A., Meyers, B.C., and Wing, R.A. (2016). Evolution of plant genome architecture. Genome Biol. **17**: 37.

Whitby, M.C. (2010). The FANCM family of DNA helicases/translocases. DNA Repair (Amst). **9**: 224–236.

White, J., and Jenkins, G. (1988). Elimination of multivalents duringmeiotic prophase in Scilla autumnalis. I. Diploid and triploid. Genome 30: 930–939.

Wijnker, E. et al. (2013). The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. Elife **2013**.

Wijnker, E., van Dun, K., de Snoo, C.B., Lelivelt, C.L.C., Keurentjes, J.J.B., Naharudin, N.S., Ravi, M., Chan, S.W.L., de Jong, H., and Dirks, R. (2012). Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. Nat. Genet. **44**: 467–470.

Wijnker, E. and de Jong, H. (2008). Managing meiotic recombination in plant breeding. Trends Plant Sci. **13**: 640–646.

Wijnker, E. and Schnittger, A. (2013). Control of the meiotic cell division program in plants. Plant Reprod. **26**: 143–158.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S., and Freeling, M. (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol. **8**: e1000409.

Wu, R., Gallo-Meagher, M., Littell, R.C., and Zeng, Z.B. (2001). A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics **159**: 869–882.

Xiong, Z., Gaeta, R.T., and Pires, J.C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus. Proc. Natl. Acad. Sci. U. S. A. **108**: 7908–7913.

Xu, Z. and Wang, H. (2007). LTR-FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**.

Yandeau-Nelson, M.D., Zhou, Q., Yao, H., Xu, X., Nikolau, B.J., and Schnable, P.S. (2005). MuDR transposase increases the frequency of meiotic crossovers in the vicinity of a Mu insertion in the maize a1 gene. Genetics **169**: 917–29.

Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of Arabidopsis male meiocytes from high-throughput sequencing: The complexity and evolution of the meiotic process. Plant J. **65**: 503–516.

Yant, L., Hollister, J.D., Wright, K.M., Arnold, B.J., Higgins, J.D., Franklin, F.C.H., and Bomblies, K. (2013). Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. Curr. Biol. **23**: 2151–2156.

Yoo, M.-J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014). Nonadditive gene expression in polyploids. Annu. Rev. Genet. **48**: 485–517.

Zamariola, L., Tiang, C.L., De Storme, N., Pawlowski, W., and Geelen, D. (2014). Chromosome segregation in plant meiosis. Front. Plant Sci. **5**: 1–19.

Zhang, L., Kim, K.P., Kleckner, N.E., and Storlazzi, A. (2011). Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. Proc. Natl. Acad. Sci. U. S. A. **108**: 20036–41.

Zhang, L., Liang, Z., Hutchinson, J., and Kleckner, N. (2014). Crossover Patterning by the Beam-Film Model: Analysis and Implications. PLoS Genet. **10**.

Zhang, Y. and Oliver, B. (2007). Dosage compensation goes global. Curr. Opin. Genet. Dev. **17**: 113–120.

Zhang, Z. et al. (2015). Genome-Wide Mapping of Structural Variations Reveals a Copy Number Variant That Determines Reproductive Morphology in Cucumber. Plant Cell **27**: 1595–604.

Zhou, A. and Pawlowski, W.P. (2014). Regulation of meiotic gene expression in plants. Plant Genet. Genomics **5**.

Zhou, R., Moshgabadi, N., and Adams, K.L. (2011). Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc. Natl. Acad. Sci. **108**: 16122–16127.

Zickler, D. and Kleckner, N. (2016). A few of our favorite things: Pairing, the bouquet, crossover interference and evolution of meiosis. Semin. Cell Dev. Biol. **54**: 135–148.

Zickler, D. and Kleckner, N. (2015). Recombination, Pairing, and Synapsisof Homologs during Meiosis. Cold Spring Harb Perspect Biol **7**: a016626.

Ziolkowski, P. a et al. (2015). Juxtaposition of heterozygosity and homozygosity during meiosis causes reciprocal crossover remodeling via interference. Elife **4**: 1–29.

Annexes

Résumé substanciel de la these en français

En amélioration des plantes, la mise sur le marché d'une nouvelle variété est l'objectif ultime du sélectionneur. Pour être adopté par l'agriculteur, les nouvelles variétés doivent présenter des traits d'intérêt agronomiques surclassant la concurrence. Développer de nouvelles variétés suppose alors de disposer d'un réservoir suffisant de diversité en amont dans lequel il est possible de puiser. Cette diversité exploitable par le sélectionneur existe à l'état naturel, que ce soit au sein d'une même espèce ou entre espèces. En terme de génétique, cette diversité correspond aux différentes versions d'un même gène (les allèles) qui sont portées par les chromosomes (support de l'information génétique). Le processus de sélection consiste alors à combiner le meilleur de la variabilité allélique existante entre individus issus d'une même espèce (croisement intraspécifique) ou entre espèces proches (croisement interspécifiques) dans une variété dite élite.

C'est au cours de la méiose que se créent les nouvelles combinaisons alléliques grâce au brassage génétique produit par la recombinaison méiotique. La méiose consiste en deux divisions cellulaires successives suivant une étape unique de réplication de l'ADN ce qui conduit mécaniquement à réduire de moitié le nombre de chromosomes. La première division permet la séparation des chromosomes dits homologues (d'origine maternelle ou paternelle), tandis que la seconde division permet la séparation des chromatides sœurs (les deux copies identiques issues de la réplication du chromosome). La recombinaison méiotique se produit au cours de la première division de méiose, elle est initiée par la formation de cassures programmées sur la molécule d'ADN. Ces cassures peuvent être réparées sous la forme d'échange réciproque, Crossing–Over (CO), ou non réciproque, Non Crossing–Over (NCO) d'information génétique entre chromosomes homologues. Ce sont les COs qui conduisent au brassage de l'information génétique.

Si la formation d'un CO dit obligatoire est nécessaire pour assurer la bonne ségrégation des chromosomes, le nombre de COs ne dépasse généralement pas 3 COs par chromosome chez la plupart des espèces. Cette régulation fine est le fait de nombreux acteurs moléculaires qui interviennent pour réguler positivement, facteurs pro-COs, ou négativement, facteurs anti-COs, le nombre de COs dans la cellule. Parmi ces derniers, on retrouve notamment la protéine FANCM, le premier régulateur négatif des fréquences de recombinaison identifié chez l'espèce modèle *Arabidopsis thaliana.* Chez le mutant *fancm*, on observe jusqu'à 3 fois plus de COs par rapport à une plante sauvage (non mutée). Malgré l'intérêt que représente le contrôle des fréquences de recombinaison pour les sélectionneurs, l'activité anti-CO de FANCM chez les espèces cultivées n'a pas été évaluée.

Chez les plantes, la polyploïdie (le fait de combiner plusieurs génomes dans une même cellule) joue un rôle majeur dans l'évolution des génomes. Parmi les plantes cultivées, de très nombreuses espèces sont des allopolyploïdes récents, c'est à dire issues du croisement de deux espèces proches suivi du doublement du stock chromosomique de l'hybride. Chez ces plantes, il est important que la formation des COs soit limitée aux seuls chromosomes dits homologues (issue du même génome), au détriment des chromosomes dits homoéologues (issue de génomes apparentés mais distincts), pour assurer une bonne ségrégation des chromosomes et la fertilité de l'espèce. Si l'on sait que de multiples mécanismes sont apparus indépendamment chez les espèces allopolyploïdes au cours de l'évolution pour inhiber la recombinaison entre homoéologues, très peu ont été décrits au niveau moléculaire. C'est le cas par exemple du locus *Ph1* chez le blé. *Ph1* correspond à un cluster de gènes intervenant dans le cycle cellulaire (cdk like). Au niveau fonctionnel, il a été montré que l'absence de *Ph1* conduit à des modifications importantes du niveau de transcription de plusieurs gènes méiotiques, que ce soient certaines des cdk situées dans les régions homéologues à *Ph1* ou le gène codant la protéine ASY1.

Le contrôle de la recombinaison entre homoéologues a aussi été très étudié chez le colza, une jeune espèce allopolyploïde (AACC, 2n=38) issue de l'hybridation du chou (CC, 2n=18) et de la navette (AA, 2n=20). Un déterminant génétique majeur, *PrBn*, intervenant dans le contrôle de la recombinaison a été identifié en utilisant des plantes allohaploïdes (AC, n=19). Chez ces plantes, les COs se forment nécessairement entre chromosomes homoéologues. L'identité de *PrBn* et des autres déterminants génétiques reste à ce jour inconnue.

Chez le colza, comme pour la plupart des espèces allopolyploides, les mécanismes limitant la recombinaison entre chromosomes homoéologues ne sont pas infaillibles. La formation résiduelle de COs entre chromosomes homoéologues résulte en la formation de larges échanges intergénomiques qui conduisent au remplacement d'une région chromosomique (qui est « perdue ») par la région correspondante portée par le chromosome homoéologue (qui est dupliquée). Ces échanges seront appelés HE par la suite. Chez le colza par exemple, on retrouve dans le génome de nombreuses HEs conduisant au remplacement d'un fragment du génome C par son équivalent sur le génome A (résumé par AACC -> AAAA).

Les grandes questions de recherche abordées au cours de ma thèse s'articulent autour de la problématique générale du contrôle des fréquences de recombinaison, que ce soit entre chromosomes homoéologues ou entre chromosomes homologues. En effet, si les acteurs moléculaires intervenant lors de la recombinaison méiotique, sont de mieux en mieux décrits, on sait peu de choses sur ce qui pourrait faire varier, en particulier à la hausse, le nombre de CO et donc les fréquences de recombinaison.

- En premier lieu, j'ai cherché à mieux caractériser ce qui gouverne les variations naturelles pour les fréquences de COs chez le colza. Pour ce faire j'ai comparé le niveau d'expression des gènes exprimés en méiose entre 2 variétés de colza, *Darmor-bzh* et *Yudal*, qui diffèrent en terme de fréquence de recombinaison entre chromosomes homoéologues (estimés chez les plantes allohaploïdes (AC, n=19), voir plus haut) et chromosomes homologues. Mon objectif était de vérifier dans quelle mesure le niveau d'expression des gènes exprimés en méiose varie entre ces 2 variétés et d'identifier les principaux facteurs responsables de cette variation.
- Dans un second temps, j'ai mis en œuvre une approche de biologie translationnelle pour vérifier que le rôle anti-CO de FANCM mis en évidence chez *Arabidopsis thaliana* est bien conservé chez le colza et la navette, 2 espèces cultivées au sein du genre *Brassica*.

Pour mener à bien le premier axe de ma thèse, j'ai travaillé sur des données de transcriptomiques (RNAseq) générées avant mon arrivée. Le matériel végétal a été extrait à partir de méiocytes (cellules en cours de méiose) de colza pour 4 conditions: 2 génotypes, *Darmor-bzh* et *Yudal* et 2 niveaux de ploïdie, euploïde (AACC) et haploïde (AC), avec pour chaque condition 3 réplicas biologiques et 2 réplicas techniques.

Le principal résultat de cette analyse est que le niveau d'expression des gènes exprimés en méiose est très variable, les différences constatées étant à la fois très répétables et pour la plupart d'entre elles de faible amplitude. Les facteurs contribuant le plus aux différences d'expression sont par ordre d'importance, le génome (A ou C), le génotype et le niveau de ploïdie.
Les différences d'expression entre les copies homéologues (portées par le génome A ou C) d'un gène est la principale source de variation observée dans notre jeu de données. Cela ne signifie pas pour autant qu'un génome domine l'autre en terme d'expression : pour un gène donné, la copie A est surexprimé par rapport à la copie C dans 35% des cas tandis que la situation opposée (C>A) est vrai dans 41% des cas. L'effet variété constitue la seconde source de variation : une majorité (60%) des gènes transcrits en méiose sont différentiellement exprimés (DE) entre *Darmor-bzh* et *Yudal*. Parmi ces gènes DE, nous avons identifié un sous ensemble de régulateurs de la transcription qui sont surexprimés dans *Yudal*. Il reste à déterminer dans quelle mesure les différences d'expressions massives observées entre génotypes sont le reflet des différences d'expression observées entre ces facteurs de transcriptions. De façon surprenante, étant donné que la recombinaison méiotique est un processus très régulé, la variation du niveau de ploïdie de la cellule (AACC vs AC) n'a eu qu'un impact limité sur l'expression des gènes exprimés en méiose.
Dès le début des analyses, il est apparu qu'une source supplémentaire de variation devait être prise en compte pour expliquer certains de nos résultats. En effet, nous avons rapidement observé que la présence de HEs (voir plus haut) dans le génome du colza avait des conséquences non négligeables sur les analyses d'expressions différentielles. Une partie importante de mon travail de thèse a donc consisté à caractériser finement le lien entre HE et expression.

Dans un premier temps, j'ai entrepris de valider au laboratoire le contenu en HEs dans nos plantes. En effet, la présence des HEs dans le génome du colza n'avait été établie jusqu'à présent que sur des bases bio-informatiques. J'ai ainsi observé une légère différence (15%) en terme de contenu en HE entre les données de référence et mes observations. J'ai confirmé la présence de 15 et 12 HEs fixées dans les génomes de *Darmor-bzh* et de *Yudal,* respectivement. Ces HEs sont préférentiellement localisées en position distale sur les chromosomes, dans des régions où les fréquences de recombinaison et la densité en gènes sont importantes. La majorité de ces HEs (22/27) conduisent à la perte de la copie portée par le génome C (concomitante avec la duplication de la copie portée par le génome A).

Ces HEs n'étant pas communes entre *Darmor-bzh* et *Yudal,* cela aboutit à une différence en terme du nombre de copie pour les gènes contenus dans les HEs (dans l'exemple précédent : AAAA dans *Darmor-bzh* contre AA dans *Yudal* et 0C dans *Darmor-bzh* contre CC dans *Yudal*). J'ai montré que cette différence en terme de nombre de copie dans les HEs se traduit par de très forte différence d'expression entre génotypes. Le contraste est tel qu'il est possible de détecter la présence de HEs uniquement sur la base des données d'expression, à la fois pour les régions perdues comme pour les régions dupliquées. Cette approche nous a d'ailleurs permis d'identifier 2 nouvelles HEs dans le génome de *Yudal*.

De façon inattendue, cette même approche nous a permis de mettre en évidence une HE nouvellement formée ségrégeant entre les 3 réplicas biologiques de *Darmor-bzh.* Cette HE aboutit à la perte de 1 (ACCC) ou 2 copies (CCCC) du génome A (comparé à la situation initiale AACC). J'ai mis à profit l'existence de ces 3 réplicas pour étudier la relation entre nombre de copie d'un gène et leur niveau d'expression. Globalement, nous avons observé que le niveau d'expression d'un gène au sein d'une HE nouvellement formée est directement proportionnel au nombre de copies de ce gène. Il y a pour la grande majorité des gènes (95%) une additivité quasi-parfaite entre copies qui conduit, dans >40% des cas, à des variations significatives du niveau d'expression total (obtenu en combinant les expressions des copies A et C). Ainsi, la duplication d'un gène faiblement exprimé (par rapport à son homéologue, qui est lui perdu) conduit à un niveau d'expression total plus faible, alors que la duplication d'un gène fortement exprimé (par rapport à son homéologue), conduit à un niveau d'expression total plus fort que chez la plante de départ AACC.

On peut s'attendre à ce que les HEs qui conduisent aux changements d'expression les plus drastiques ne soient pas neutres pour la plante. Un certain nombre d'indices suggèrent que ces événements ne sont pas conservés tel quel dans le génome et que leurs impacts tendent à être atténués au fil du temps. En comparant l'expression des gènes dans les HEs récentes (HEs entre replicas biologiques de *Darmor-bzh)* et celles plus anciennes (HEs fixées entre *Darmor-bzh* et *Yudal*), j'ai notamment observé une sous représentation des gènes dont la duplication conduit à une diminution du niveau total d'expression dans les HEs anciennes. J'ai aussi observé pour ces gènes que le niveau d'expression était plus important qu'attendu, le niveau d'expression total se rapprochant ainsi du niveau observé dans la situation pré-HE (AACC). Ces observations laissent supposer que des mécanismes de compensation, via une régulation de l'expression ou une sélection s'exerçant contre certains gènes au sein des HEs, se sont progressivement mis en œuvre pour limiter l'impact des HEs sur l'expression.

La stratégie de biologie translationnelle que j'ai mis en œuvre dans le 2e volet de ma thèse repose sur l'obtention de mutants pour chacune des copies de FANCM présentes chez le colza et la navette. FANCM est présent en 1 copie chez la navette (*BraA.FANCM*) et en 2 copies chez le colza, une sur le génome A et une sur le génome C *(BnaA.FANCM* et *BnaC.FANCM,* respectivement). Le faible nombre de copie de FANCM chez le colza et la navette a permis la mise en œuvre d'une approche dite de TILLING pour rechercher des mutants au sein de populations mutagénéisés à l'aide d'un agent chimique (EMS). Parmi les nombreuses mutations identifiées, j'ai retenu celles prédites pour affecter le plus sévèrement l'activité anti-CO de FANCM. Pour la navette j'ai choisi 1 mutations conduisant à la substitution d'un acide aminé très conservé dans un domaine nécessaire à l'activité anti-CO de FANCM (*BraA.fancm-1*). Pour le colza, j'ai retenu pour la copie A, une mutation STOP conduisant à une version tronquée de la protéine (*BnaA.fancm-1*) et pour la copie C, en l'absence de STOP, j'ai retenu 2 mutations conduisant à la substitution d'acides aminés très conservés (*BnaC.fancm-1* et *BnaC.fancm-2*).

Chez la navette, j'ai mis en œuvre une approche cytologique pour montrer que *Bra.fancm-1* est capable de restaurer la formation de CO chez un mutant déficient pour les fréquences de recombinaison (*braA.msh4-1*$^{-/-}$). Ainsi j'ai observé 2 fois plus de COs chez le double mutant (*braA.msh4-1*$^{-/-}$ *braA.fancm-1*$^{-/-}$) comparé au simple mutant **(***braA.msh4-1*$^{-/-}$). Cela montre que *BraA.FANCM* limite les fréquences de recombinaisons chez la navette.

Chez le colza, j'ai analysé l'effet des mutations *fancm* en comparant les fréquences de recombinaison entre des plantes sœurs mutantes et sauvages pour les 2 copies (A et C) de FANCM. J'ai observé une augmentation modeste (~30%) mais répétable des fréquences de CO chez l'un des deux double mutants que j'ai analysé (*bnaA.fancm-1*$^{-/-}$*bnaC.fancm-2*$^{-/-}$). Cette augmentation concerne à la fois la formation des COs entre chromosomes homologues (estimé chez le colza euploïde, AACC) et entre chromosomes homoéologues (estimé chez des allohaploides de colza, AC).

Si ces résultats montrent bien que FANCM est un régulateur négatif de la recombinaison chez 2 plantes cultivées au sein du genre *Brassica*, ils ne permettent pas de savoir dans quelle mesure FANCM limite les fréquences de recombinaison chez le colza. J'ai en effet pu montrer que la mutation BnaC.fancm-2 chez *Arabidopsis thaliana* n'est pas une mutation perte de fonction. Par conséquent, une activité résiduelle de BnaC.fancm-2 est attendue chez le colza.

université PARIS-SACLAY | ÉCOLE DOCTORALE Sciences du végétal : du gène à l'écosystème

Titre : Vers une caractérisation fonctionnelle de la recombinaison méiotique chez le colza ; analyse du transcriptome méiotique et de mutants hyper-recombinants.

Mots clés : Allopolyploïde, transcriptome, variants structuraux, recombinaison méioitque, FANCM

Résumé : La recombinaison méiotique produite par les Crossing Overs (COs) est un facteur limitant pour l'efficacité de la sélection variétale. Une possibilité pour produire des plantes hyper-recombinantes serait d'exploiter la variabilité intraspécifique pour les fréquences de recombinaison. L'identification des polymorphismes causaux, liés à la séquence ou l'expression, représente un travail de longue haleine. Une approche alternative serait de produire des mutants pour des régulateurs négatifs des fréquences de recombinaison. Chez le colza, jeune allotétraploïde (AACC, 2n=38), il est possible de jouer sur ces 2 approches. Dans un premier temps j'ai cherché à vérifier dans quelle mesure pouvait varier le transcriptome méiotique entre 2 variétés ayant servi à cartographier un QTL pour le contrôle de la recombinaison entre chromosome homoéologues (hérités des génomes parentaux). Ce transcriptome méiotique s'est révélé de façon inattendu très variable ; les principales sources de cette variation étant notamment la nature du génome (A ou C) ainsi que l'effet variété.

J'ai montré que les HEs (le remplacement d'une région chromosomique par la duplication de la région homoéologue) contribuent de façon importante aux différences d'expression observées à la fois entre variétés ou au sein d'un même génotype. Dans un second temps, j'ai vérifié que FANCM décrit chez *Arabidopis thaliana* comme un régulateur négatif pour les fréquences de recombinaison avait bien la même fonction chez les *Brassica*. Chez *Brassica rapa* j'ai vérifié qu'un mutant *fancm* complémente comme attendu un mutant déficient pour la voie majoritaire de formation des COs. Chez *Brassica napus* j'ai observé une faible augmentation à la fois des fréquences de recombinaison entre chromosomes homologues et homoéologues. Ce travail souligne l'importance de la caractérisation des HEs chez les allopolyploïdes. Au delà de leurs impacts sur le contenu et l'expression génique, les HEs ont très certainement des conséquences phénotypiques. Cette étude présente aussi un exemple de biologie translationnelle pour un trait important en amélioration des plantes.

Title: Towards a functional characterization of meiotic recombination in rapeseed; analysis of the meiotic transcriptome and hyper-recombinant mutants

Keywords: Allopolyploid, transcriptome, homoeologous exchange, meiotic recombination, FANCM

Abstract: Meiotic recombination driven by Crossing-Over (CO) is a limiting factor for the efficiency of plant breeding. One way to produce hyper-recombinant plants is to use the existing interspecific variability for recombination frequencies. Identification of the causal polymorphisms, either link to gene sequence or expression, represents a long-term endeavour. Another possibility is to mutate anti-meiotic CO genes. In rapeseed, a young allotetraploid species (AACC, 2n=38), both of these approaches are possible.

First I wanted to check how much varies the meiotic transcriptome between 2 varieties that differ in term of recombination between homoeologous chromosomes (inherited from parental genomes). Unexpectedly, the meiotic transcriptome turned out to be very variable, the main source of this variation being notably the origin of the genome (A or C) and the variety.

I also showed that homoeologous exchanges (HEs; the replacement of one chromosomal region with a duplicate of the homoeologous region) contributed to this variation and led to large changes in expression both between and within varieties. Then I assessed whether FANCM, an anti-CO protein identified in *Arabidopis thaliana* had the same function in the *Brassica* genus. In *Brassica rapa*, a *fancm* mutant complements as expected a meiosis mutant defective in the main formation pathway for the formation of meiotic COs. In *Brassica napus*, I observed a slight increase in both homologous and homoeologous recombination frequencies. This work emphasizes the importance of characterizing HEs in allopolyploids species. Beyond their impact on gene content and expression, HEs most have likely phenotypic consequences. This study also presents an example of translational biology for an important trait in crop breeding.